TN-00.4041.251

INTRODUCTORY TRANSISTOR THEORY

T.L. Schiesel

IBM Technical Note

Publications Department
Product Development Laboratory
International Business Machines Corporation
Poughkeepsie, New York

DATE: February 20, 1958
REVISED: April 28, 1958

TN 004.041.251

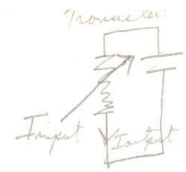
INTRODUCTORY TRANSISTOR THEORY

by

T. L. Schiesel

ABSTRACT

"Introductory Transistor Theory" presents a non-mathematical approach to transistor physics and is intended to provide the fundamental concepts needed to use these devices. Included is a discussion of atomic structures, N and P type semiconductors, current carriers, junction diodes, forward and reverse currents, breakdown, junction transistors, and basic transistor connections. The material contained in this article has been compiled from lecture notes given by the author.



INTRODUCTORY TRANSISTOR THEORY

The discovery of the transistor in 1948 and its rapid development since that time, has opened up an entirely new field of electronics. In the few short years of their existance, transistors have come into their own by taking their place along side vacuum tubes in radio, television, computers and many widespread experimental projects. In some respects, the operation of the transistor may be compared to a vacuum tube; however, the action within the device is quite different. In the vacuum tube, output current is controlled by a small input voltage varying an electrostatic field at the grid. In the transistor, a varying input current controls the amount of output current. The internal resistance is controlled by the input signal and through this action the transistor gets its name; it is a TRANSfer resISTOR.

To gain an acceptable understanding of transistor action, it becomes necessary to investigate atomic and electron theory in more detail than was necessary for an equivalent understanding of vacuum tubes. For this reason, considerable emphasis is placed on atomic structures and behaviors. This approach may appear too sophisticated for a purely practical approach, but remember, the explanations presented would seem over-simplified to the scientist. The intent is to give you an intuitive understanding of how this device operates so that you won't have to regard a transistor as a mysterious "black box."

In some ways, the transistor can be considered a mere infant, because much of its theory and many of its applications are still being developed. Several radically new concepts are involved; do not be misled into believing this is just another version of a standard electronic device. The study of transistors should be approached with an open mind and a willingness to accept theories and ideas that are foreign to conventional vacuum tube electronics.

Transistors are ideal components for computers because of their small size, low power consumption, long life, and extreme flexibility in circuit design. Transistors are divided into several basic types each having its own peculiarities and advantages. The alloy junction type is described throughout this manual, and is becoming more widely used, because its characteristics include rugged construction, low noise level, good stability, low operating voltages, good power handling ability and efficiency and a reasonably good frequency response.

The basic material in transistors is germanium. This element is extracted in its crude form as a by-product of zinc refining. Processing the ore results in white germanium dioxide powders. A silver -grey ingot of germanium is obtained by heating the powder to about 650° in an atmosphere of hydrogen gas. For further purification, a section of the ingot is heated by an RF induction coil and the ingot is moved as the section becomes molten. In effect, the molten zone moves along the ingot. Impurities prefer the molten to the solid state; therefore, they can be "drawn off" to one end as the ingot is moved. This is called a "zone refining" process because one area or zone of the material is purified as it is heated. The degree of purity can be determined by its electrical resistivity which increases as impurities are removed. Commercially pure germanium has a resistivity of 47 ohms per cubic centimeter at 25° C. An impurity ratio of only one part in one hundred million reduces the resistivity to 4 ohms per cubic centimeter! The impurity of germanium cannot be less than the 4 ohm per cubic centimeter measurement. The refined metal is called "intrinsic" germaium.

ATOMIC STRUCTURES

The study of transistor fundamentals is dependent upon a basic understanding of the composition of matter and the structure of the smallest part of an element, the atom. Transistors are solid state devices containing atoms of germanium arranged in a definite geometric pattern. Current flow through this device depends upon electron movement through the solid. A thorough understanding of this type of transfer can be gained only by analyzing typical atomic structures. A basic knowledge of atomic theory may be sufficient to comprehend transistor action; however, a rather detailed explanation is included for those who desire a review for more complete background of this subject.

An atomic structure is subdivided into a positively charged core called the nucleus and a cloud of negatively charged electrons. The electrons revolve about the nucleus in orbits that are not necessarily circular or in the same plane, but there is a certain average discreet orbit in which any electron tends to stay. These electron orbits can be considered analogus to the planetary orbits of the solar system. For simplicity, these orbits are normally pictured as being circular and in the same plane.

A hydrogen atom (Figure 1) is an example of an atomic structure in its simplest form; it contains only one electron revolving around a positively charged nucleus. The negative charge of the electron is exactly equal to the positive charge of the nucleus, therefore, the atom is electrically neutral. Other elements have atomic structures that are more complex than this simple example; however, the only difference lies in the number and arrangement of the electrons and the positive charge of the nucleus. All atoms are electrically neutral because the number of negative charges always equals the number of positive charges. As the structure of the atom becomes more complex, the positive charge of the nucleus increases and the number of orbital electrons also increases to maintain the electrical balance.

The basic material in transistors is germanium. This element is extracted in its crude form as a by-product of zinc refining. Processing the ore results in white germanium dioxide powders. A silver -grey ingot of germanium is obtained by heating the powder to about 650° in an atmosphere of hydrogen gas. For further purification, a section of the ingot is heated by an RF induction coil and the ingot is moved as the section becomes molten. In effect, the molten zone moves along the ingot. Impurities prefer the molten to the solid state; therefore, they can be "drawn off" to one end as the ingot is moved. This is called a "zone refining" process because one area or zone of the material is purified as it is heated. The degree of purity can be determined by its electrical resistivity which increases as impurities are removed. Commercially pure germanium has a resistivity of 47 ohms per cubic centimeter at 25° C. An impurity ratio of only one part in one hundred million reduces the resistivity to 4 ohms per cubic centimeter! The impurity of germanium cannot be less than the 4 ohm per cubic centimeter measurement. The refined metal is called "intrinsic" germaium.

ATOMIC STRUCTURES

The study of transistor fundamentals is dependent upon a basic understanding of the composition of matter and the structure of the smallest part of an element, the atom. Transistors are solid state devices containing atoms of germanium arranged in a definite geometric pattern. Current flow through this device depends upon electron movement through the solid. A thorough understanding of this type of transfer can be gained only by analyzing typical atomic structures. A basic knowledge of atomic theory may be sufficient to comprehend transistor action; however, a rather detailed explanation is included for those who desire a review for more complete background of this subject.

An atomic structure is subdivided into a positively charged core called the nucleus and a cloud of negatively charged electrons. The electrons revolve about the nucleus in orbits that are not necessarily circular or in the same plane, but there is a certain average discreet orbit in which any electron tends to stay. These electron orbits can be considered analogus to the planetary orbits of the solar system. For simplicity, these orbits are normally pictured as being circular and in the same plane.

A hydrogen atom (Figure 1) is an example of an atomic structure in its simplest form; it contains only one electron revolving around a positively charged nucleus. The negative charge of the electron is exactly equal to the positive charge of the nucleus, therefore, the atom is electrically neutral. Other elements have atomic structures that are more complex than this simple example; however, the only difference lies in the number and arrangement of the electrons and the positive charge of the nucleus. All atoms are electrically neutral because the number of negative charges always equals the number of positive charges. As the structure of the atom becomes more complex, the positive charge of the nucleus increases and the number of orbital electrons also increases to maintain the electrical balance.

Hydrogen Atom
Figure 1

An electron is held in its orbit by two balanced forces acting upon it. The positively charged nucleus tends to attract the electron to the center but this is counteracted by the centrifugal force tending to propel the electrons away from the nucleus. Attraction to the nucleus increases in proportion to the positive nuclear charge; however, the attractive force diminishes rapidly as the electron revolves in orbits further from the atom center.

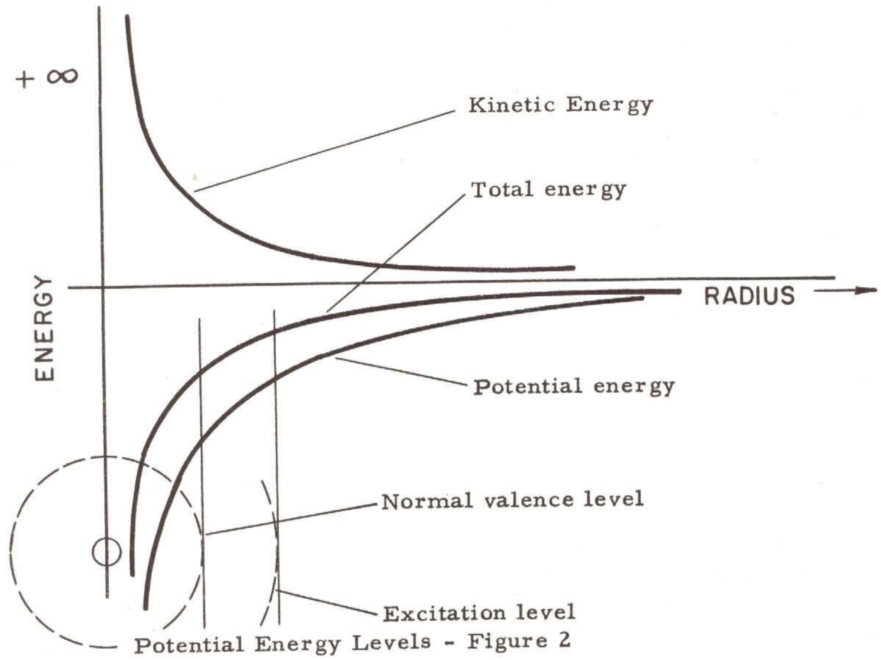
The energy of bodies in motion is called kinetic energy; the amount of energy is a function of mass and velocity. The formula for kinetic energy ($E = 1/2\ mv^2$) indicates that velocity has a greater influence than the mass on the amount of energy generated by the moving body. Electrons rotating in large orbits travel at a lower velocity than electrons in smaller orbits. This lower velocity decreases the kinetic energy of electrons revolving in orbits of increased diameters.

Potential Energy Level (Figure 2)

Energy may also appear in forms other than kinetic energy. Heat, light, electric, or potential are all forms of energy that may be inherent in a given element. These latter forms of energy appear quite different but they are related and each can be converted into any of the other forms. An example of this transformation is the development of heat and light when electrical energy is supplied to a light bulb.

Potential energy is an energy that may be stored, awaiting release. Potential energy of an electron is twice the value of its kinetic energy but is opposite in sign; therefore, the potential energy of the electron increases as the electron orbit becomes larger. The position or level of the electron orbit is, therefore, related to the kinetic and potential energies of the electron. The level at which an electron may exist is called the potential energy level. Highest potential energy levels exist in orbits furthest from the nucleus.

In a stable atom, an electron can leave its orbit only by gaining or losing potential energy. When an electron gains potential energy, its orbit becomes larger because it can overcome, to a greater degree, the attractive force of the nucleus. If potential energy is lost, the orbit becomes smaller until the two acting forces are again in balance.



Potential Energy Levels - Figure 2

Energy Bands (Figure 3)

A fundamental concept, called the Pauli exclusion principle, states that no two electrons in an atomic structure can occupy the same energy level. Atoms containing many electrons have energy levels, in which electrons can establish orbits, at various distances from the nucleus. These levels do not radiate in a uniform pattern about the atom center but rather tend to group themselves together to form bands of energy. Between each band is an energy gap; this is a forbidden region for an energy level. Only certain numbers of energy levels can be contained in each band. Each band of energy, starting with the innermost band is assigned a letter code to distinguish it from other bands. The letters used are K, L, M, N, O, P, and Q.

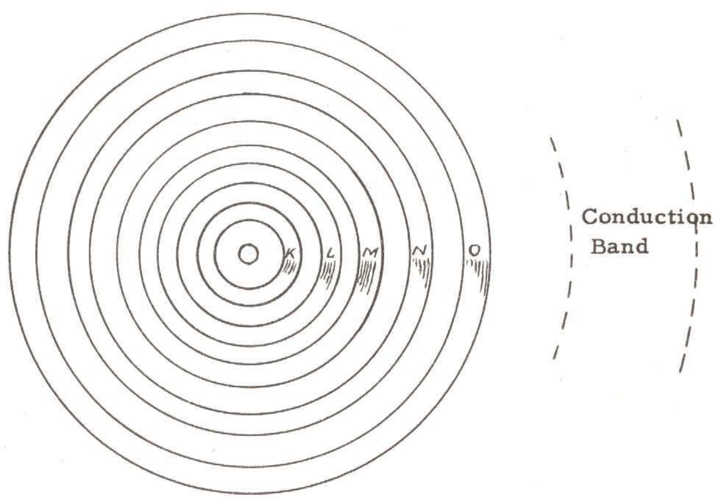
The first (K) band can contain only $2 \times 1^2 = 2$ energy levels.
The second (L) band can contain only $2 \times 2^2 = 8$ energy levels.
The third (M) band can contain only $2 \times 3^2 = 18$ energy levels.
The fourth (N) band can contain only $2 \times 4^2 = 32$ energy levels.

Although the following bands are never completely filled, their capacities are presumed as follows:

The fifth (O) band can contain only $2 \times 5^2 = 50$ energy levels.
The sixth (P) band can contain only $2 \times 6^2 = 72$ energy levels.
The seventh (Q) band can contain only $2 \times 7^2 = 98$ energy levels.

It has been stated that as atomic structures become more complex, increased numbers of electrons are rotating in the orbital energy levels. The high attractive force of the nucleus makes it safe to assume that electrons occupy the inner orbits first then tend to find places in the outer orbits. The inner orbit electrons, therefore, arrange themselves in bands of 2, 8, and 18 electrons. These electrons remain in their orbits regardless of most applied outside influences and are called the "filled bands" of the atom.

The outermost band (regardless of which band is involved) contains the electrons that go into chemical and electrical combinations with other atoms. These are the "valence electrons" of the atom and exist in the "valence band" of energy. Valence band energy levels are higher than those of the filled band because the electrons are rotating in orbits further from the nucleus. When an outside influence is applied it is possible to cause an electron to acquire sufficient energy to jump over the forbidden region into the next higher level called the conduction band. Once an electron is in the conduction band it is free of any influence of the nucleus and then becomes a carrier of electric current. It is important to note that electrons never land in the forbidden region between the valence and conduction levels; they either transfer to the conduction band or remain in the valence band.



Atomic Energy Bands
Figure 3

Table of Elements

Elements		(2) K	(8) L	(18) M	(32) N	(50) O
1 HYDROGEN	H	1				
2 HELIUM	He	2				
3 LITHIUM	Li	"	1			
4 BERYLLIUM	Be	"	2			
5 BORON	B	"	3			
6 CARBON	C	"	4			
7 NITROGEN	N	"	5			
8 OXYGEN	O	"	6			
9 FLUORINE	F	"	7			
10 NEON	Ne	"	8			
11 SODIUM	Na	"	"	1		
12 MAGNESIUM	Mg	"	"	2		
13 ALUMINUM	Al	"	"	3		
14 SILICON	Si	"	"	4		
15 PHOSPHORUS	P	"	"	5		
16 SULFUR	S	"	"	6		
17 CHOLORINE	Cl	"	"	7		
18 ARGON	A	"	"	8		
19 POTASSIUM	K	"	"	8	1	
20 CALCIUM	Ca	"	"	8	2	
21 SCANDIUM	Sc	"	"	9	2	
22 TITANIUM	Ti	"	"	10	2	
23 VANADIUM	V	"	"	11	2	
24 CHROMIUM	Cr	"	"	13	1	
25 MANGANESE	Mn	"	"	13	2	
26 IRON	Fe	"	"	14	2	
27 COBALT	Co	"	"	15	2	
28 NICKEL	Ni	"	"	16	2	
29 COPPER	Cu	"	"	18	1	
30 ZINC	Zn	"	"	"	2	
31 GALLIUM	Ga	"	"	"	3	
32 GERMANIUM	Ge	"	"	"	4	
33 ARSENIC	As	"	"	"	5	
34 SELENIUM	Se	"	"	"	6	
35 BROMINE	Br	"	"	"	7	
36 KRYPTON	Kr	"	"	"	8	
37 RUBIDIUM	Rb	"	"	"	8	1
38 STRONTIUM	Sr	"	"	"	8	2
39 YTTRIUM	Y	"	"	"	9	2
40 ZIRCONIUM	Zr	"	"	"	10	2
41 NIOBIUM	Nb	"	"	"	12	1
42 MOLYBDENUM	Mo	"	"	"	13	1
43 TECHNETIUM	Tc	"	"	"	13	2
44 RUTHENIUM	Ru	"	"	"	15	1
45 RHODIUM	Rh	"	"	"	16	1
46 PALLADIUM	Pd	"	"	"	18	
47 SILVER	Ag	"	"	"	18	1
48 CADMIUM	Cd	"	"	"	18	2
49 INDIUM	In	"	"	"	18	3
50 TIN	Sn	"	"	"	18	4
51 ANTIMONY	Sb	"	"	"	18	5

		(2) K	(8) L	(18) M	(32) N	(50) O	(72) P	(98) Q
52 TELLURIUM	Te	2	8	18	18	6		
53 IODINE	I	"	"	"	"	7		
54 XENON	Xe	"	"	"	"	8		
55 CESIUM	Cs	"	"	"	"	8	1	
56 BARIUM	Ba	"	"	"	"	8	2	
57 LANTHANUM	La	"	"	"	18	9	"	
58 CERIUM	Ce	"	"	"	20	8	"	
59 PRASEODYMIUM	Pr	"	"	"	21	8	"	
60 NEODYMIUM	Nd	"	"	"	22	8	"	
61 PROMETHIUM	Pm	"	"	"	23	8	"	
62 SAMARIUM	Sm	"	"	"	24	8	"	
63 EUROPIUM	Eu	"	"	"	25	8	"	
64 GADOLINIUM	Gd	"	"	"	25	9	"	
65 TERBIUM	Tb	"	"	"	26	9	"	
66 DYSPROSIUM	Dy	"	"	"	28	8	"	
67 HOLMIUM	Ho	"	"	"	29	8	"	
68 ERBIUM	Er	"	"	"	30	8	"	
69 THULIUM	Tm	"	"	"	31	8	"	
70 YTTERBIUM	Yb	"	"	"	32	8	"	
71 LUTETIUM	Lu	"	"	"	32	9	"	
72 HAFNIUM	Hf	"	"	"	"	10	"	
73 TANTALUM	Ta	"	"	"	"	11	"	
74 TUNGSTEN	W	"	"	"	"	12	"	
75 RHENIUM	Re	"	"	"	"	13	"	
76 OSMIUM	Os	"	"	"	"	14	"	
77 IRIDIUM	Ir	"	"	"	"	15	"	
78 PLATINUM	Pt	"	"	"	"	17	1	
79 GOLD	Au	"	"	"	"	18	1	
80 MERCURY	Hg	"	"	"	"	"	2	
81 THALLIUM	Tl	"	"	"	"	"	3	
82 LEAD	Pb	"	"	"	"	"	4	
83 BISMUTH	Bi	"	"	"	"	"	5	
84 POLONIUM	Po	"	"	"	"	"	6	
85 ASTATINE	At	"	"	"	"	"	7	
86 RADON	Rn	"	"	"	"	"	8	
87 FRANCIUM	Fr	"	"	"	"	"	8	1
88 RADIUM	Ra	"	"	"	"	"	8	2
89 ACTINIUM	Ac	"	"	"	"	"	9	"
90 THORIUM	Th	"	"	"	"	"	10	"
91 PROTACTINIUM	Pa	"	"	"	"	20	9	"
92 URANIUM	U	"	"	"	"	21	9	"
93 NEPTUNIUM	Np	"	"	"	"	23	8	"
94 PLUTONIUM	Pu	"	"	"	"	24	8	"
95 AMERICIUM	Am	"	"	"	"	25	8	"
96 CURIUM	Cm	"	"	"	"	25	9	"
97 BERKELIUM	Bk	"	"	"	"	25	10	"
98 CALIFORNIUM	Cf	"	"	"	"	27	9	"
99 EINSTEINIUM	E	"	"	"	"	28	"	"
100 FERMIUM	Fm	"	"	"	"	29	"	"
101 MENDELEVIUM	Md	"	"	"	"	30	"	"
102 NOBELIUM	No	"	"	"	"	31	"	"

Figure 4

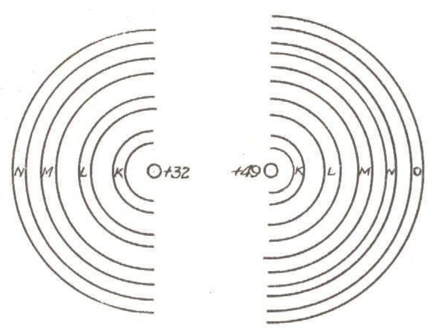
Arrangement of Electrons in Energy Levels (Figure 4)

The table showing the arrangement of electrons in the energy levels of atoms shows that the K band contains one electron for the hydrogen atom and contains no more than 2 for all the remaining elements. As soon as a band of energy levels is filled, electrons must then seek orbits in the next band further from the nucleus. As the progression of elements becomes more complex, the energy levels in the L band fill up until the band contains 8 electrons for the element neon; this is the maximum number of electrons that can be contained in this band. Further progressions of the elements now force electrons to establish orbits in the M band which has 18 possible energy levels. Electrons increase in the outer orbits until a maximum of eight is reached; at this point, the element becomes absolutely inactive chemically and cannot combine with any other element. This "Octet" theory of eight electrons in the outer orbit is exhibited in the inert gasses Neon, Argon, Krypton, Xenon, and Radon. As soon as eight electrons are contained in an outer orbit, further electron orbits must be established in the next outer band regardless of whether empty energy levels remain in that band. The stable condition demonstrated by this octet theory becomes one of the fundamentals on which transistor materials can be analyzed.

Energy Level Relationships Between Different Elements (Figure 5)

Electron orbits are governed by the velocity of the electron and the nuclear charge of the atom center. The increased positive charge of the more complex atoms develops a greater attractive force for electrons and resultant orbits are established nearer the nucleus. It has already been established that electrons traveling in orbits near the nucleus have a smaller potential energy than those electrons traveling in more remote orbits. From these facts we can surmize that similar bands (K, L, M, N, etc) are at different energy levels for different elements. An example of this is shown for two of the more common elements used in transistors. The K, L, M, and N bands are closest to the nucleus of the indium atom because the nucleus contains a more positive charge than the germanium atom.

Germanium



Indium

Figure 5 Energy Level Relationships

Energy Gaps (Figure 6)

Differences in the energy levels between elements contribute to its electrical as well as its chemical behavior. Elements with many electrons in their outer shell tend to approach the stable condition existing in the inert gasses by taking on electrons until the valence band octet is complete. In these elements, a wide energy gap exists between the valence band and the conduction band; it is difficult for electrons to acquire sufficient energy to jump into the conduction band; however, it is quite easy for an electron to "fall" into the empty energy level. A basic concept of electricity states that to have an electric current there must be carriers of electric current. Valence electrons become the electric carriers when they are excited to the conduction band. If electrons do not reach the conduction band there is no transfer of current and the element displays a high resistance to the flow of current. This atomic arrangement exists in insulators such as mica or glass and is shown in Figure 6; the valence and conduction bands are separated by a wide energy gap. Energy gaps are measured in electron-volts which is the amount of energy gained or lost when electrons are accelerated through a potential difference of 1 volt. For a reference of comparison, a typical insulator may have an energy gap of 1 electron-volt.

Elements with few electrons in their outer orbits also tend to become stable by giving up their outer orbit electrons. This results in outer orbits that again contain eight electrons and the element approaches the characteristics of a stable element. In the atomic structures of elements with only a few electrons in the outer orbits, the gap between the valence and conduction bands is small, or does not exist at all. It is very easy for these elements to have valence electrons excited to the conduction band and thereby become the free carriers of electric current. Certain, such as silver or copper contain only one electron in their outer orbits, therefore; are good conductors. The energy gap may be in the order of .05 electron volts. If a sufficient amount of energy is applied, electrons can be forced to leave the metal. This property is used in the cathode of a vacuum tube and the movement of electrons through the vacuum constitutes a current flow.

When the number of electrons in the valence band lies between the extremes of an insulator and a conductor the width of the energy gap also adjusts itself to an intermediate value. Germanium satisfies the requirements of this intermediate range by containing four electrons in its valence orbits. Elements possessing this type of structure are called semi-conductors because only limited numbers of valence electrons can be excited to the conduction band. At room temperature, germanium absorbs enough energy so that some electrons are able to jump from the valence band to the conduction band. The energy gap for germanium is .7 electron-volts. The comparisons between the energy gaps of various types of elements should make it apparent why the resistivities of semi-conductors is higher than that of metalic conductors but less than that of insulators.

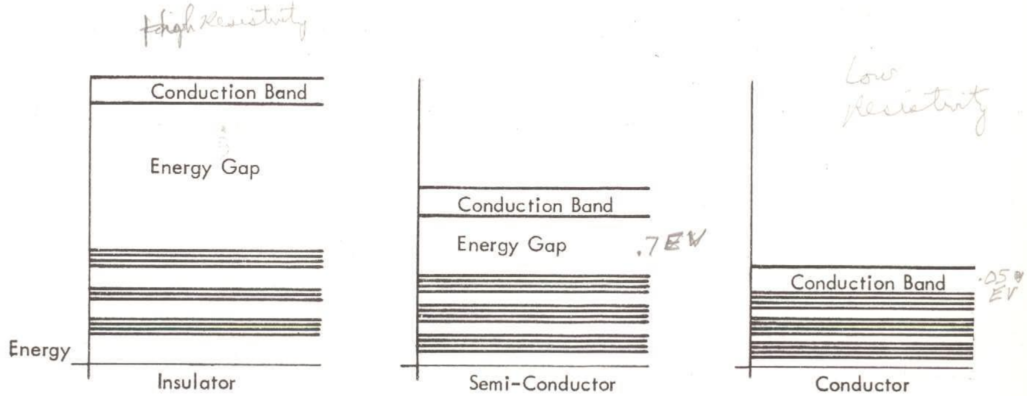


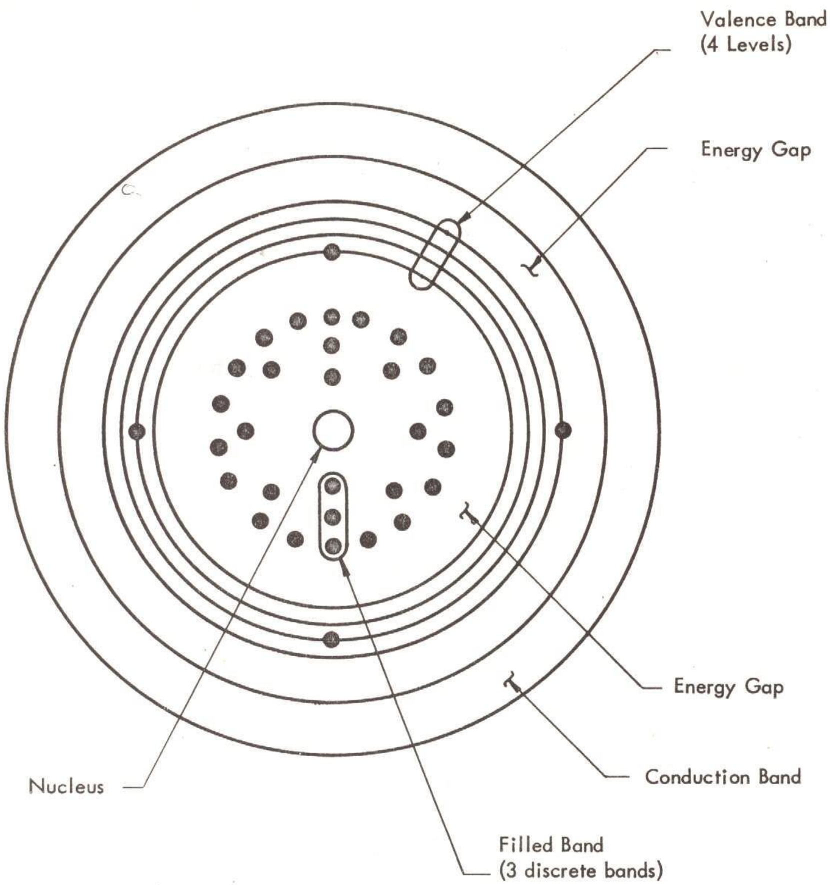
Figure 6 Energy Gap Relationships

Germanium Atomic Structure (Figure 7)

Several different types of semi-conductor materials are used in transistor manufacture; however, germanium, silicon arsenic, antimony and indium are common elements found in junction devices. Germanium is a basic element: therefore, the description of its structure is given the most detail.

The Germanium nucleus contains 32 plus charges and is surrounded by 32 electrons revolving in four bands containing 2, 8, 18, and 4 electrons respectively. Normally, the atom is electrically neutral; however, when the atom contains more or less than the normal number of electrons it possesses a net charge and is called an "ion".

Because an external source of energy must be added to an electron before it can leave its orbit for a higher energy level (outer orbit), it is assumed that an electron may give up some of its energy to jump to a lower level. When an electron changes its energy level frequently it is in an "unstable" state; if the probability of change is small, the electron is in a "stable" state. Inner orbit electrons are the most stable because they exist at a lower energy level. Outer orbit electrons are rather unstable and are more easily removed; therefore, when atoms are ionized, the ionization is restricted to the unfilled rings. For these reasons, chemical compositions involve interaction among electrons only in the unfilled (valence) rings. The filled ring electrons have little or no influence on chemical or electrical action; therefore, only the valence electrons need be discussed in further analysis.

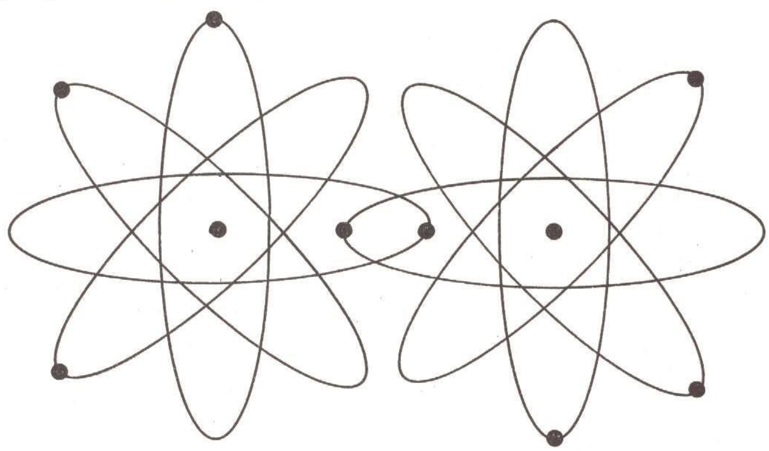


Schematic Diagram of Germanium Atom

Figure 7

Germanium Molecule (Figure 8)

If two identical atoms are brought together to form a molecule, the valence, or highest, energy levels are shared by both atoms in the molecule. This means that electrons in the outer orbits are not localized to either atom but have orbits allowing them to travel throughout the molecule and serve to bind the atoms together. The unfilled levels lying above the valence band are called the excitation, or conduction, levels of the molecule. These levels may contain electrons for brief periods of time when valence electrons are raised in energy by the absorption of light energy, by bombardment, from thermal motion of the atom center or by high electric fields.



Two Atom Molecule
Figure 8

Co-Valent Bonding (Figure 9)

Valence electrons rotate in orbits that are far removed from the nucleus; in fact, valence orbits of one atom cross the valence orbits of its neighbor. During the course of its travel, an electron is acted upon by its nucleus and the nuclei of its neighboring atoms. In addition to this action, there is an attractive force set up between electrons in adjoining valence orbits. Each electron spins on its axis and develops a magnetic field, in much the same way as the earth spins on its axis and develops a magnetic field as it travels around the sun. The magnetic field of one electron becomes additive to the magnetic field set up by the valence electrons of adjoining atoms. This electron pairing is called "co-valent" bonding. The potential energy of the paired electrons is effectively lowered because the bonding process makes it difficult for the valence electrons to become free carriers.

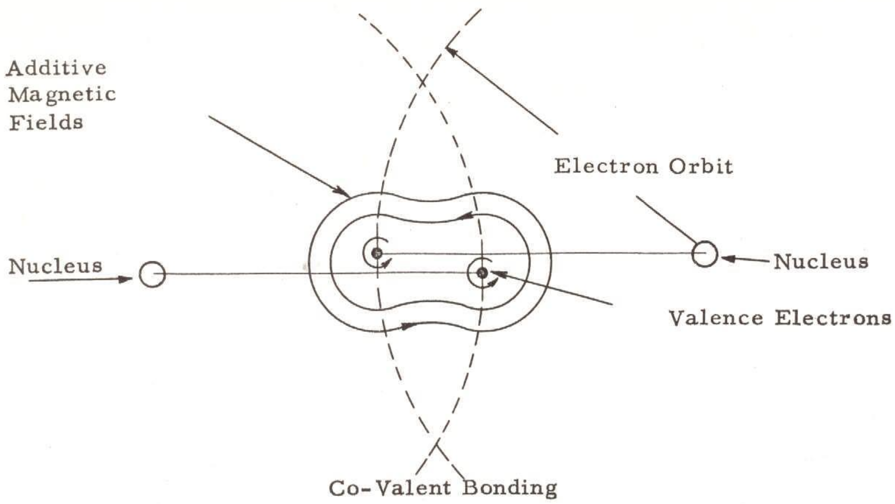
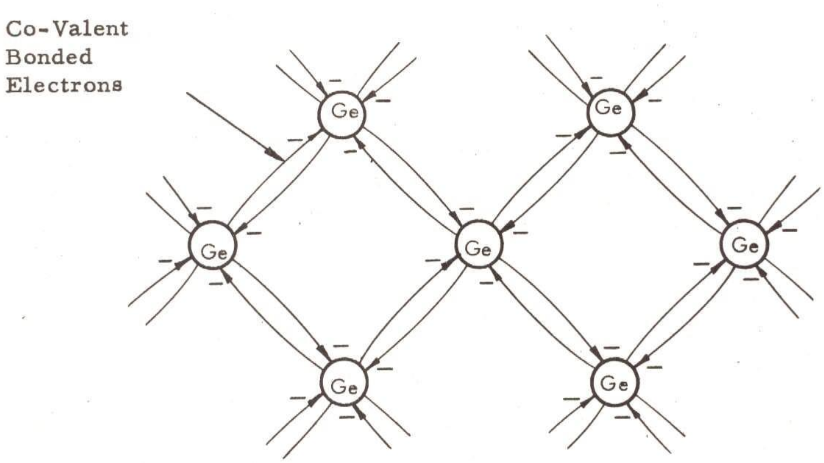


Figure 9

The co-valent bonding of the valence electrons stabilizes the germanium atoms because each atom effectively has eight electrons in its outer orbit. The balance of energy existing between adjacent valence electrons also determines the geometric arrangement of the atoms. This arrangement of atoms is called a "lattice". The co-valent bonding of the four valence electrons of germanium develops a diamond shaped lattice represented in the single plane drawing of Figure 10. Each electron appearing in an electron-pair co-valent bond is more rigidly bound to a fixed orbit than are unbonded electrons. An important comparison can be made at this point. About fifteen times as much energy is needed to free a co-valent bonded electrons as is required to free an unbonded valence electron.

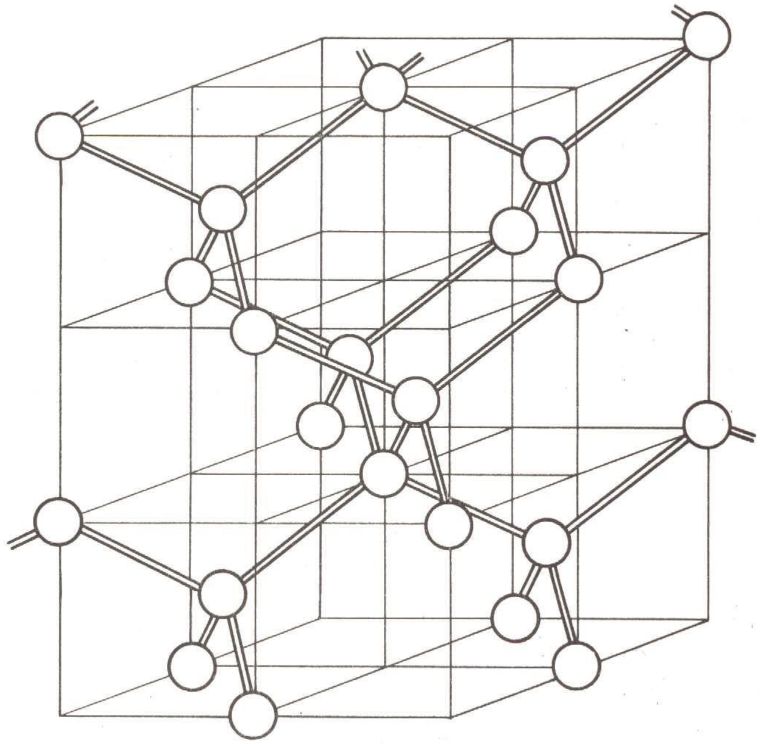


Diamond Shaped Germanium Crystal Lattice

Figure 10

Germanium Crystal Structure (Figure 11)

Bonding continues between atoms until a point of stability is reached where the valence band of each atom effectively contains eight electrons. This co-valent electron pairing binds the atoms together so that a solidified sample of pure germanium forms a three dimentional crystal whose cubical lattice is as illustrated. Single crystal formations are essential to transistor operation. The balls in the diagram represent the atom centers and the **heavy** connecting lines denote the co-valent electrons.



Germanium Crystal Lattice Showing Bonds Between Adjacent Atoms

Figure 11

Arsenic

Another element used in transistors is Arsenic. Its atomic structure contains bands of 2, 8, 18 and 5 electrons. The significant point to remember is that this element contains one more electron than germanium in its outer orbit.

Indium

The third element used in transistors is Indium. Its atomic structure contains bands of 2, 8, 18, 18, and 3 electrons. The significant point to remember is that this element contains one less electron than germanium in its outer orbit.

Temperature Coefficients (Figure 12)

The resistivity of conductors increases with increased temperature; this is called a positive temperature coefficient. Conductors possess many electrons that are available for the conduction process; the number of these electrons cannot appreciably increase with an increase in temperature. Elevated temperatures cause atoms to vibrate about their equilibrium positions with larger and larger amplitudes. These vibrating atoms cause collisions and a scattering of current carrying electrons. Scattering results in a net reduction in the velocity of the electrons and effectively reduces current flow through the conductor. Reduced current flow, and increased resistance are the net results of increased temperatures acting upon materials possessing a positive temperature coefficient.

The resistivity of semiconductors decreases with increased temperature; this is called a negative temperature coefficient. An energy gap exists between the valence and conduction bands; therefore, the number of electrons in the conduction band is a function of the energy applied to the valence electrons. When the applied energy is in the form of heat, the number of electrons available for the conduction process increases with increased temperatures. The lattice structure of the crystal is altered by vibrating atoms, and the mean free electron path is decreased by elevated temperatures just as in conductors. This effect is overcome by the rapidly increasing number of electrons elevated to the conduction band. The number of electrons rises exponentally with temperature and resistivity decreases to indicate a negative temperature coefficient.

Chemically pure germanium has a rather high resistivity and is a relatively poor conductor; therefore, a carefully controlled amount of impurity must be added to the metal before it can be used in a transistor. Impurities are added at the rate of about one part in 10 million. The type of impurity added becomes the determining factor in the operation of the final product.

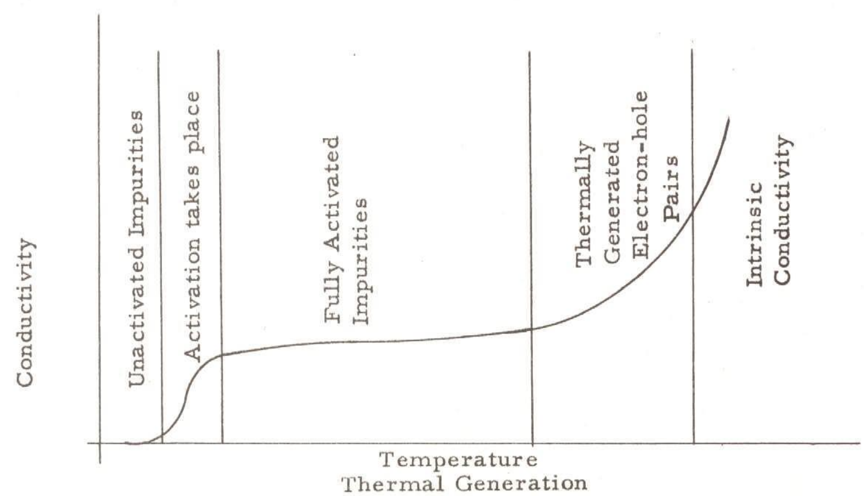


Figure 12

Current Carriers - Electrons and Holes

From the discussion of temperature coefficients it can be seen that electrons can gain enough thermal energy to break their co-valent bonds and become free carriers. Once they have been liberated from the influence of the nucleus, they wander throughout the crystal in no apparent direction. A significant departure from conventional electron carrier theories develops at this point. The atom losing a valence electron, due to thermal agitation, possesses a vacant energy level. This empty level is actually an electron Deficiency and is called a "hole" because it may be regarded as if a physical hole appeared in the atom.

Germanium crystals develop many free electrons, and consequently many holes, by the thermal agitation taking place at normal room temperature. Free electrons wander around aimlessly if there is no particular place for them to go; however, holes left in other atoms present an attractive force to wandering electrons. Holes are filled by these free electrons and the atom again becomes electrically neutral.

Assume an electron becomes free and leaves behind a hole. Now assume an atom to its left also liberates an electron and this carrier fills the originally generated hole. The net effect of this transfer is that an electron moved to the right and a hole moved to the left. The directions of these two charges aid, rather than cancel, the total amount of current developed by the transfer.

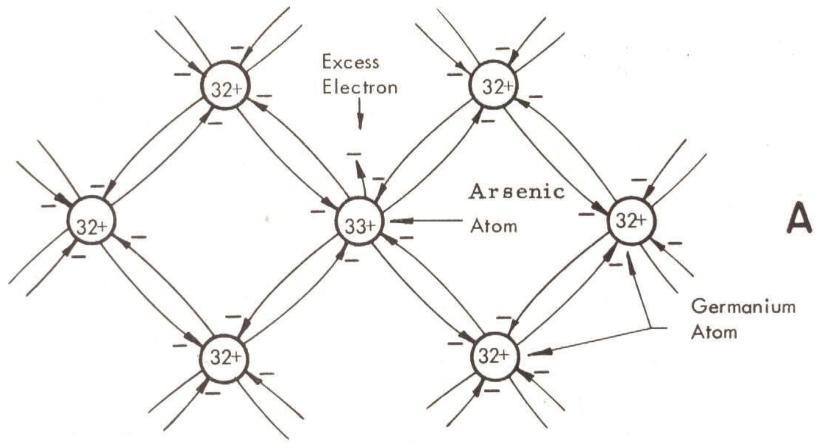
The term "hole" is used as an aid to expand on the accepted electron theory; it becomes effective in explaining the transfer of charges throughout a crystal. The basic theory of electron flow must be expanded to include this flow of holes before transistor action can be appreciated. The hole is considered to have a positive charge and the transfer of holes is in a direction opposite to the direction of electron flow.

Diffusion

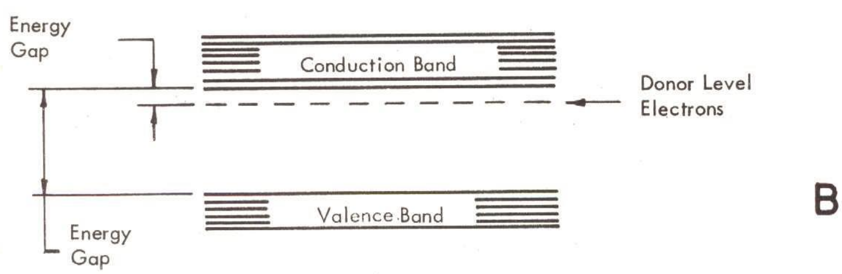
A process known as diffusion results when a solid contains an unequal concentration of carriers throughout its mass. Assume a sample of germanium develops a high concentration of electrons at one end. Due to the concentration, and the random motion of electrons, the electrons spread out or "diffuse" until there is an equal concentration of electrons throughout the crystal. This spreading occurs without the application of a voltage and is called "diffusion current".

Drift

The random motion of thermally agitated electrons is altered by the application of an electric potential. Electrons travel toward the positive potential and the current in this directed path is called "drift current". The net displacement of holes and electrons is effected by diffusion and drift. Holes drift in a direction opposite to electrons because they carry an opposite charge.



N Type Germanium Crystal



Energy Level of Donor Electrons

N Type Germanium

Figure 13

N Type Germanium (Figure 13)

Arsenic is a typical impurity that is added to germanium; its valence band contains 5 electrons and therefore is called a "penta-valent element. This impurity enters the germanium lattice when small amounts are added to the intrinsic material. Four of its outer orbit electrons band with the germanium valence electrons to form a stable lattice of germanium and a arsenic atoms. The extra valence electron, which is not in a co-valent bond, is at an energy level just below the conduction band of the germanium structure. Germanium doped with penta-valent impurities is called N-type because the majority of the conducting carriers are negative.

Majority Carrier Current - Electrons

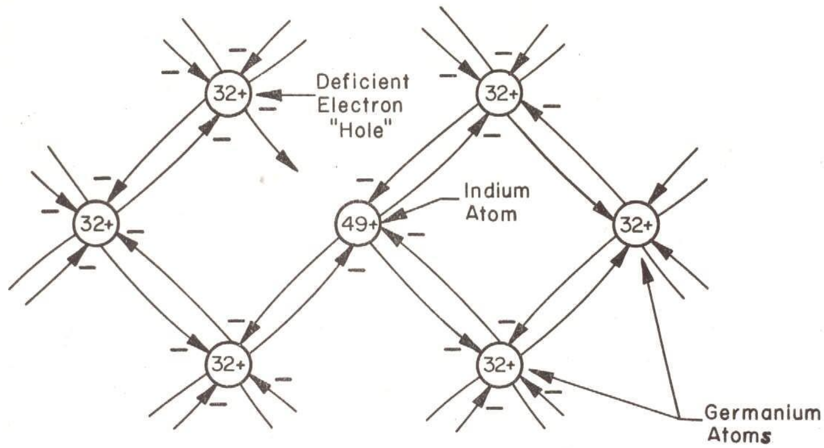
Conduction in N-type germanium is by means of "donor" electrons because the excess, unbonded electrons are easily donated to the conduction band.

The energy level of donor electrons is called the donor level and the energy gap between the donor level and conduction band is much less than the energy gap between the valence band and conduction band. For this reason, donor atoms can be thermally agitated to give up their excess electrons at a much lower temperature than would be required to excite valence band electrons to the conduction band. Actually, temperatures above a few degrees centigrade are sufficient to cause donor level electrons to enter the conduction band as current carriers. When a donor level electron leaves the impurity atom, it leaves behind a positive ion. This ion is a bound positive charge because its position is fixed in the crystaline structure and it is not free to take part in a conduction process. Do not confuse this positive ion with a hole.

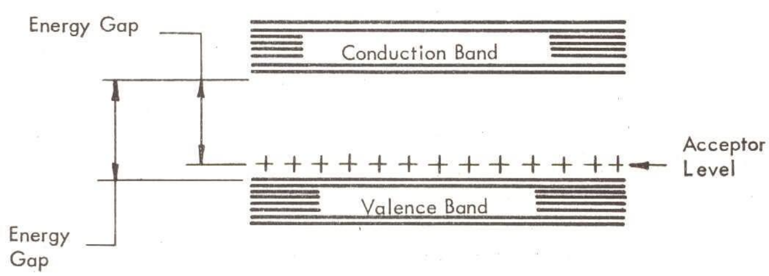
In the low to intermediate temperature ranges, donor levels give up their electrons to the conduction band and the valence band remains practically filled. In this range, an almost constant number of electrons is available for the conduction process and the N type germanium is in its positive temperature coefficient range.

Minority Carrier Current - Holes

In N-type material, higher temperatures develop electron-hole pairs just as in intrinsic germanium. The free electron is elevated to the conduction band and adds to the normal supply of majority carriers; the semiconductor now displays a negative temperature coefficient. A hole also develops in the valence band of a germanium atom. At normal operating temperature, the movement of holes is less likely to occur than free majority carrier electron current; therefore, holes are referred to as minority carriers in N-type germanium. A hole may be eliminated if a free electron falls into the lower potential energy level presented by the hole. This electron-hole recombination takes place shortly after the minority carrier is generated; consequently, minority carriers have a finite average lifetime. Holes that are not eliminated by recombination, travel toward an applied negative potential because they are mobile and can take part in an electrical conduction process.



P Type Germanium Crystal



Energy Level of Acceptor Holes

P Type Germanium

Figure 14

P Type Germanium (Figure 14)

Intrinsic germanium can also be doped with indium which has three electrons in its valence band. The resulting alloy has a lattice structure made up of germanium and indium atoms. In the description of energy level differences between various elements, it was noted that corresponding bands came closer to the nucleus as the complexity of the atom and the nuclear charge increased. Electrons in these closer bands exist at lower potential energies; however, all the energy levels need not be filled. The indium atom contains one band of electrons more than germanium; the energy level of its outer band is just slightly above the energy level of the germanium valence electrons.

The valence orbits of the indium atom effectively contain six electrons because its three electrons are in a co-valent bond with electrons of only three of its germanium neighbors. In an attempt to complete its valence octet, and because empty energy levels are available, the indium atom is ready to accept an additional electron. The vacant indium energy level is called an acceptor level because it is ready to accept an electron from any source.

Majority Carrier Current - Holes

The vacant energy levels, generated by the addition of acceptor impurities, correspond to incomplete co-valent bonding between specific atoms. This incomplete bonding makes it relatively easy for an external source of energy to cause an unbonded germanium electron to move into the unfilled level of an impurity atom. At normal operating temperatures, there is enough thermal activity to keep these.acceptor levels filled and the outer band of the indium atom contains seven electrons when this transfer takes place. The acceptor impurity receiving the electron becomes ionized and we now have a bound negative charge. The germanium atom that lost the valence electron now possesses a vacant energy level in its valence band and this is again called a hole. Identical hole flow reasoning applies to germanium "doped" with indium as applied to intrinsic germanium because the electron vacancy again appears in the valence band. The number of holes developed is equal to the number of acceptor impurity atoms in the crystal. Holes allow a current flow because it makes electrons available to the conduction process by removing the restrictions of co-valent bonding.

When an electrical potential is applied across a sample of P material, the holes migrate toward the negative terminal. Actually, electrons can be said to move from the valence band of one atom to the valence band of the next, without requiring the additional energy needed to jump into the conduction band, but here again, let us call it "hole flow." If the voltage and temperature were low, and the sample contained only P type impurities, this would be the only type of conduction to occur.

Minority Carrier Current - Electrons

In P-type material, further thermal agitation will develop electron-hole pairs just as in intrinsic germanium. The free electron is elevated to the conduction band and may drift into the hole created in the previously mentioned germanium atom so as to complete the octet of bonded electrons. This action removes the effect of the originally generated hole but establishes a hole in some other section of the crystal. The total number of holes has not changed but we have noticed that the generation of a free electron may be for only a short duration of time before it recombines with a hole and is cancelled. Electrons that do not recombine with holes become the "minority carriers" in P-type material and as normal, will move toward the positive terminal of an applied potential. Minority carrier electrons increase rapidly as temperature increases; this action makes transistors quite temperature sensitive.

P Type and N Type Potential Energy Relationships

Orbits of impurity electrons are established either closer to the nucleus, or further from the nucleus, compared to the orbital dimensions of the germanium atom. Adding Indium, with a nuclear positive charge of 49, indicates that orbital electrons would be rotating closer to the atom center. However, Indium contains one band more than germanium, which results in valence electrons rotating in orbits which are slightly larger than the germanium valence orbits. This condition establishes the potential energy level of P type valence electrons at a higher level than the valence electrons of intrinsic germanium.

In a similar manner, the potential energy level relationship can be investigated for N type materials. Adding arsenic, with a nuclear positive charge of 33, indicates that the orbital electrons would be rotating in slightly smaller orbits than the electrons of germanium atoms. Arsenic and germanium each contain the same number of electron bands.

These altered conditions establish the potential energy levels of valence and conduction bands higher in P materials than in N materials. The difference in levels is controlled by the amount of impurities that are added to the intrinsic germanium. As you might suspect, this difference increases as impurities are added. This difference in potential energy levels is an important factor that contributes to the action of semi-conductor devices.

N-P Junction Diode (Figure 15)

When two germanium alloys containing N and P type impurities are alloyed together to form a single crystal, an important action takes place at the junction between the two layers. To simplify the discussion, assume that all the holes are in the P material and all the free electrons are in the N material. This assumption neglects the effect of minority carriers that may be produced as thermally generated couples.

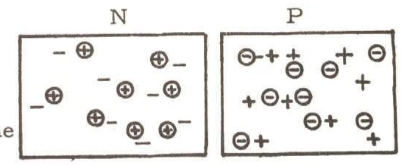
The N region contains a number of donor atoms which may be assumed to be distributed uniformly throughout the volume. At normal temperatures, most of these atoms will have lost their fifth electron, and will, therefore; possess a "bound" positive charge. The free electrons will be moving at random and will be uniformly distributed throughout the volume (minus signs in Figure 15-A). This means that the N region contains equal numbers of bound positive charges and free negative charges, and there will be no resultant electric field within the volume as yet.

Similar conditions exist in the P region. Most of the acceptor atoms will have received electrons from adjoining germanium atoms, due to thermal excitation, and will become bound negative charges. Free holes created as a result of these acceptor atoms will be diffusing at random throughout the volume. Once again, a balance exists between the positive and negative charges and there is no resultant electric field.

The N-P junction presents a situation where the number of free electrons to the left of the junction greatly outnumber the free electrons to the right of the junction (and vice versa for the holes). Random motion of free electrons on the left causes them to diffuse to an area of lower concentration. In a like manner, holes also diffuse from an area of high concentration to an area of low concentration. Considerable discussion has been given to the higher potential energy level of the P material holes. This higher level is an added influence for electrons to migrate across the junction.

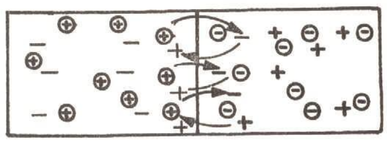
The diffusion of carriers must also be viewed from the standpoint of charged particles. When a free electron moves from the N-region it leaves behind a bound positive charge and adds one negative charge to the P-region. Adding electrons to the P material effectively adds minority carriers which create electron-hole pairs and a cancelling of charges. Conversly, when a hole moves from the P-region it leaves behind a bound negative charge and adds one positive charge to the N-region. The added holes are minority carriers in the N material and they combine with free electrons to effect an additional reduction of the negative potential of the N side. The diffusion process builds up a charge density and a corresponding electrostatic potential on either side of the junction.

free electron
bound positive Ion
number of free electrons equals the number of bound + Ions

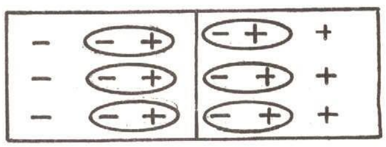


free holes
bound negative Ion
number of free holes equals number of bound - Ions

Diffusion of free carriers to area of least concentration

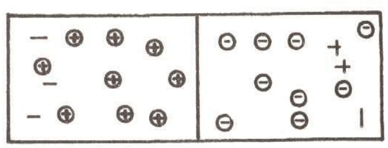


Free electrons & holes recombine in the P & N regions

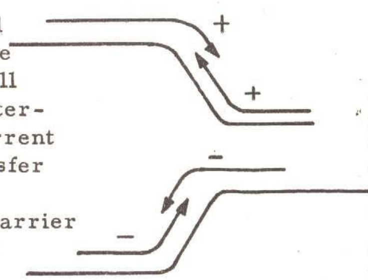


(Bound charges not shown)

Number of bound charges has not changed;therefore, N side now appears more positive than the P side



Thermally generated (minority) holes slide down the potential hill but are exactly counteracted by reverse current due to majority transfer



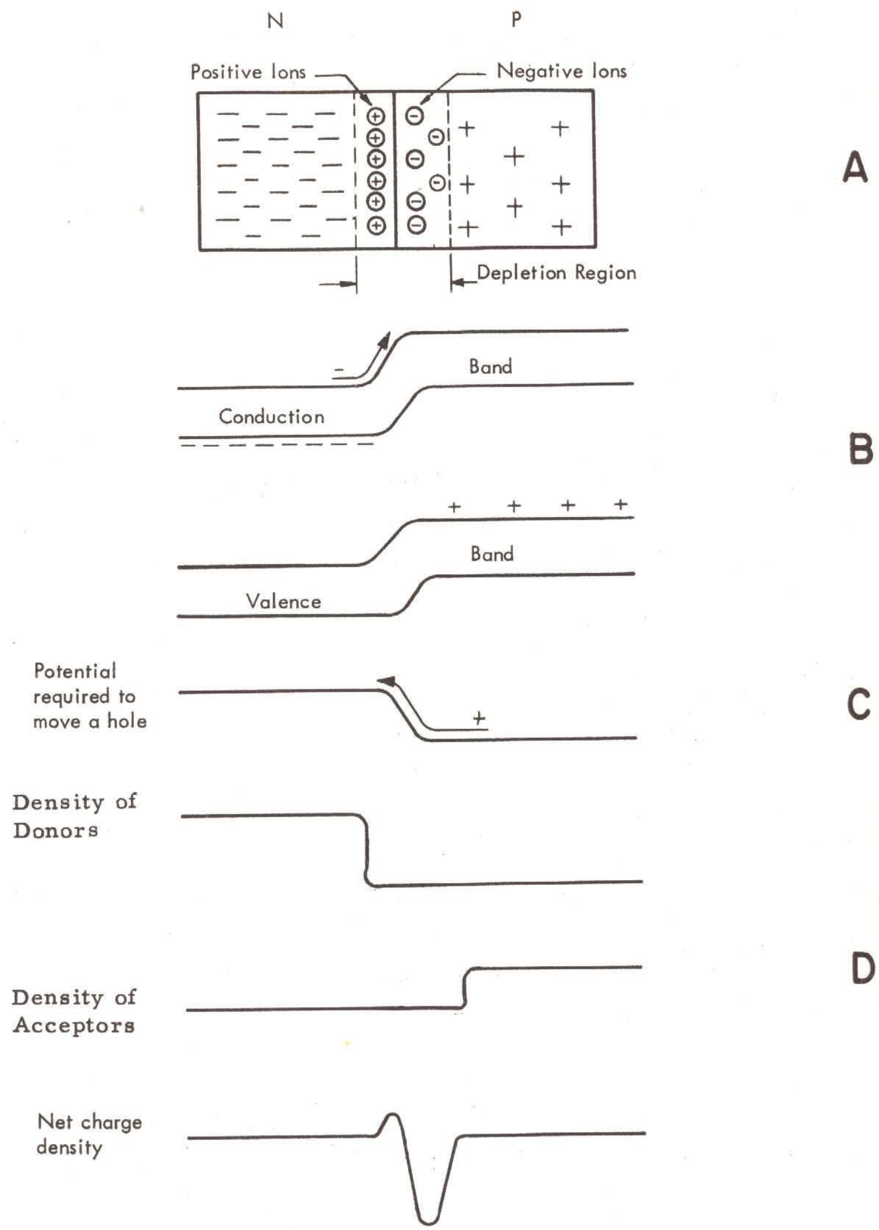
Energy a majority carrier hole must attain before it can cross the junction

Energy a majority carrier electron must attain before it can cross the junction

Thermally generated (minority) electrons slide down the potential hill but are exactly counter balanced by electrons that attain enough thermal energy to climb the potential barrier.

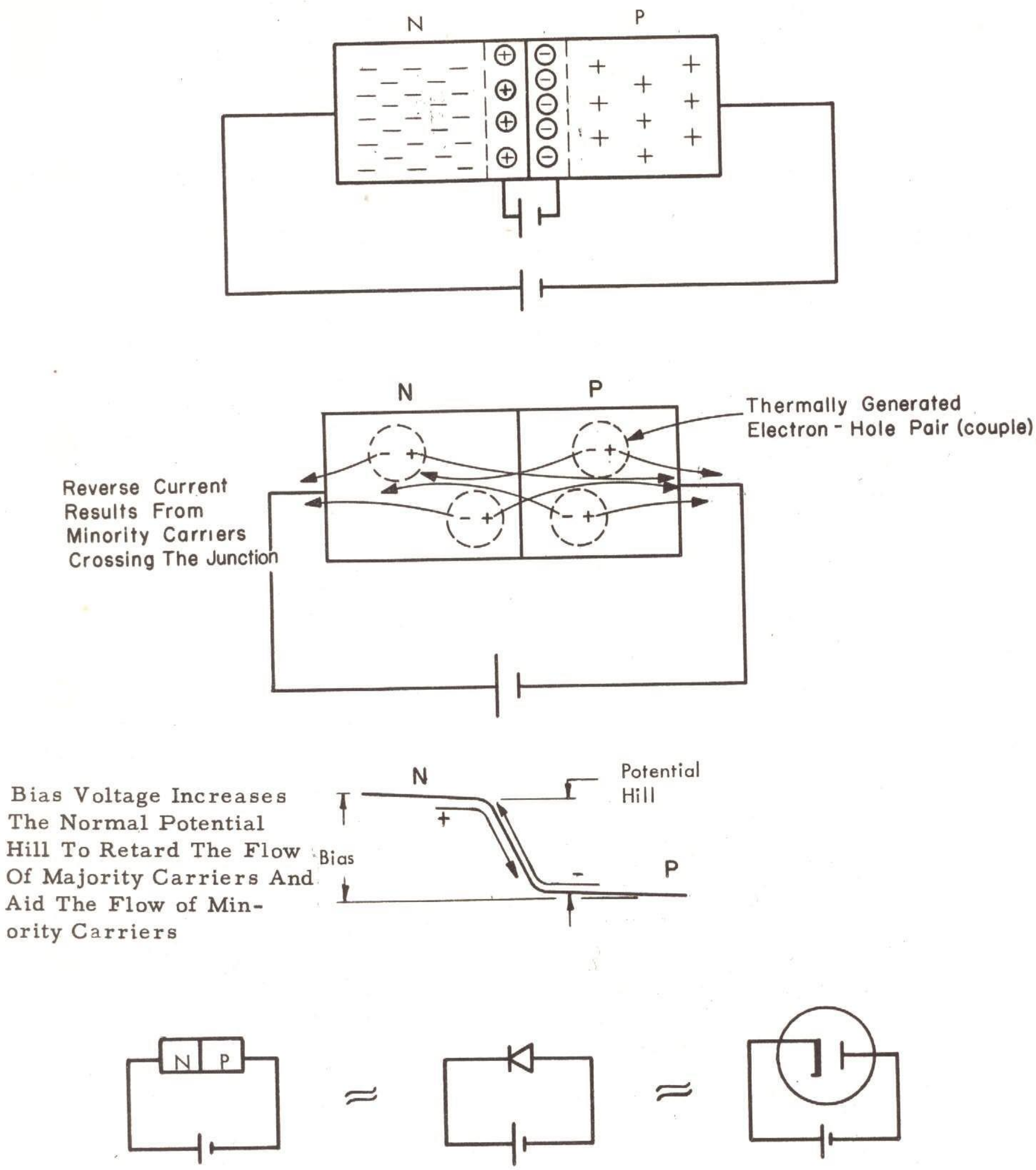
N-P JUNCTION DIODE USING EQUAL IMPURITY CONCENTRATIONS

Figure 15



NP Junction Using Unequal Impurity Concentrations

Figure 16



Reverse Biased Semiconductor Diode

Figure 17

This electrostatic potential is in effect a potential barrier or potential hill which opposes any further action of the diffusion process which initially established the barrier. In order for electrons to diffuse from the N-material to the P-material they must attain sufficient energy from an external source to overcome the opposition of the potential hill.

The average time required for a minority carrier to meet and combine with a majority carrier is considered the minority carrier lifetime. These lifetimes vary under different conditions from a fraction of a microsecond to milliseconds. The combining of carriers and the build-up of the potential hill occurs near the junction because it is a remote probability that a minority carrier will travel very far during its lifetime.

From the discussion to this point, it can be concluded that the height of the potential hill is a function of the ratio of the impurity concentrations in the two dissimilar types of material.

Figure 16 shows the conditions, existing when there is an unequal ratio of concentrations on either side of the junction.

Reverse Bias Diode (Figure 17)

The character of the barrier changes as soon as an external potential is applied. If the positive terminal of the source voltage is connected to the N-region, and the negative to the P-region, the diode is said to be reverse biased.

Let us consider only extrinsic conduction for the moment: that is, conduction of impurity carriers only. The positive terminal tends to attract free electrons from the N-region and the negative terminal tends to attract free holes from the P-region. This will cause both types of free carriers to move away from the junction leaving behind more bound charges that are not balanced. Obviously this cannot continue because there is no inexhaustable supply of free carriers to maintain this flow. In fact, when the battery is first connected, this very process immediately enlarges the depletion region near the junction. Enlarging the depletion region raises the potential hill by the amount of the reverse bias. This makes it more difficult for current carriers to attain sufficient energy to climb the potential hill and thus reach the opposite side. Because the transfer of carriers across the junction is limited, only a small current flows. The reverse biased junction is compared to a cut-off crystal or vacuum tube diode in figure 17.

Normal reverse current is developed in the range of operating temperatures because thermal agitation may cause germanium electrons to break loose from their valence bonds. Thermal generation of an electron-hole pair in the P-material creates a free, minority carrier, electron. This carrier diffuses at random and may reach the junction during its lifetime. The potential hill is in a direction to aid the transfer of electrons from the P to the N region: therefore, this minority carrier electron crosses the junction. The N region now contains one more electron than it needs for

equilibrium; the P region contains one excess hole. To maintain equilibrium, the N region delivers an electron to the positive terminal and the P region accepts an electron leaving the negative terminal of the battery. The electron entering the P region combines with an excess hole and is then neutralized. The net effect is an electron leaving the battery, flowing through the diode from P to N and entering the positive terminal of the battery.

A similar analysis can be made if a thermally agitated atom develops a free electron - hole pair in the N region. In each case the current is carried by minority carriers and not the majority carriers resulting from impurity atoms. When heat produces minority carriers a finite percentage reach and cross the junction; the remainder combine with majority carriers and are eliminated. The percentage of electrons that reach the junction depends on the density of the majority carriers, minority carrier lifetime, the physical structure of the diode, and partially on the applied bias. The number of minority carriers produced per second is a function of the temperature at the junction. This is actually an undesirable effect because this reverse current doubles for about each 10° centigrade rise.

We have said that the percentage of thermally generated carriers crossing the junction is only slightly influenced by the applied voltage. After an initially small reverse bias is applied, a further increase in bias results in practically no increase in reverse current. This reverse current remains essentially constant over a wide range of reverse bias voltage until another source of carriers is developed. A sufficiently large increase in reverse bias can develop two possible sources of additional carriers. In each case an increase in reverse current represents a breakdown of the reverse resistance of the diode. Neither type of breakdown will in itself damage the diode.

Avalanche Breakdown

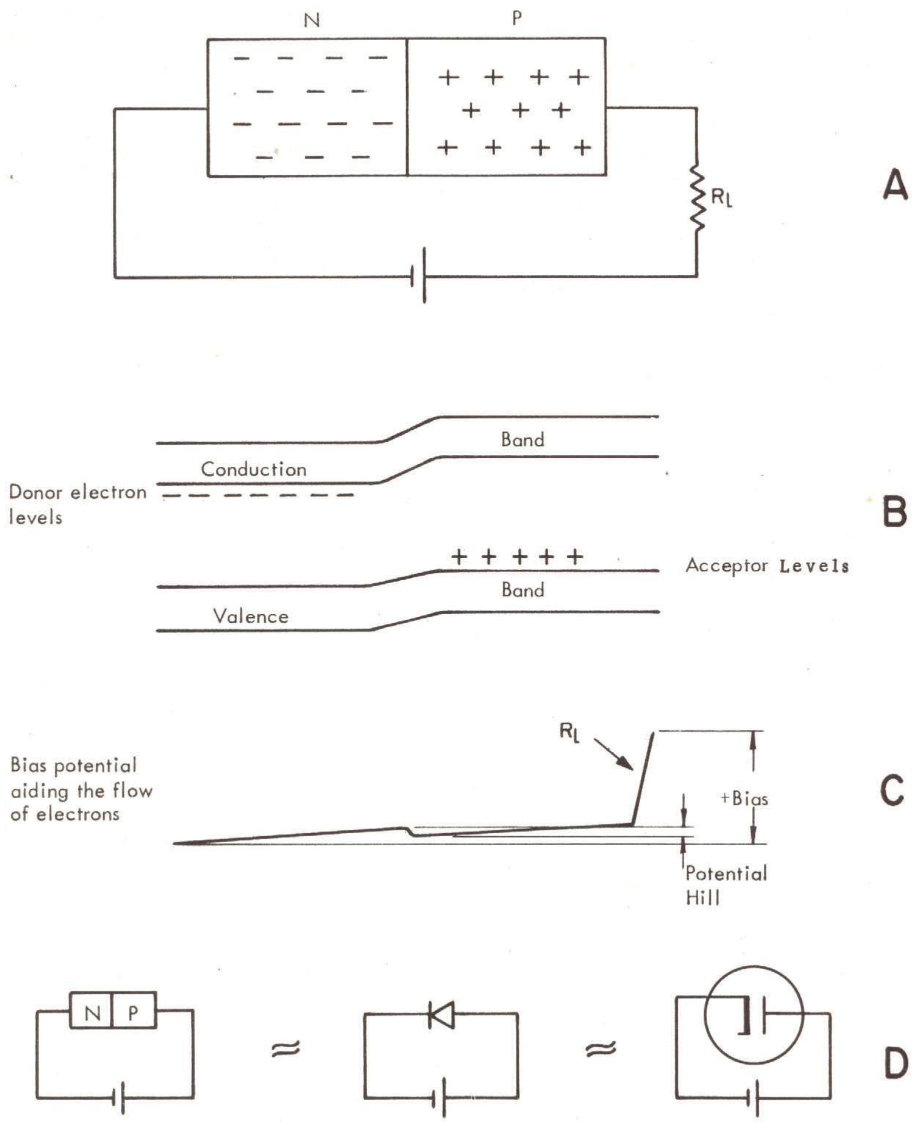
Electrons gain energy as they are moved under the influence of an electric field. As reverse bias is increased, minority carriers cross the junction and acquire sufficient energy to bombard and dislodge valence ring electrons from germanium atoms. This increases the number of electrons available in the N region that can be supplied to the positive bias return. Positive ions or minority carriers are developed in the N region that can cross the junction and repeat the process in the P region. When sufficient bias voltage is supplied to start this process, reverse current increases rapidly in a manner similar to a rolling stone starting an avalanche on a mountainside.

Zener Breakdown

This type of breakdown is named after the man who first proposed an explanation of the process. The force which an electric field exerts on an electron is proportional to the strength of the field; this is measured in

volts per centimeter. The voltage appearing across a junction may be small but it is concentrated in such a short distance that fields having a magnitude of 100,000 volts per centimeter or more may exist there. When the field approaches this strength, valence bond electrons are pulled from their orbits and provide a source of carriers.

In each breakdown case, the amount of reverse current is limited by the resistance of the external circuit and not by the diode. Breakdown currents do not damage the diode unless the power dissipation is increased to the point where temperature rise causes a physical breakdown of the junction. Power dissipation is equal to the product of the voltage after breakdown and the current flowing through the diode. A typical value of power limitations is fifty milliwatts for computer type transistors but this may be increased to several watts by suitable design of case construction, size of the junction, or the use of a heat sink.



Forward Biased Semiconductor Diode

Figure 18

Forward Biased Diode (Figure 18)

When a bias voltage is connected with the positive terminal to the P region, and the negative terminal to the N region, the diode is forward biased. As you might suspect, the applied potential repels the majority carriers of each region toward the junction. Some N material electrons may recombine with the bound positive charges at the N side of the barrier, and some P material holes may recombine with the bound negative charges at the P side of the barrier; this effects a further reduction of the potential hill at the junction. Effectively, there is a diffusion through the barrier to become minority carriers on the opposite side.

For example: Some majority carrier electrons may combine with bound positive charges but the greater percentage diffuse through to the P material to become minority carriers. Again, some of these electrons may combine with the majority carrier holes but the greater percentage are swept through the P material under the influence of the positive potential.

The high velocity of the minority carrier electrons may dislodge other valence electrons in the P material so that more carriers are added to the conduction process. The charge of the minority carrier electrons also develops an electric field that further assists the transfer of thermally generated majority carriers toward the junction. This process is cumulative and results in a flow of carriers that greatly exceeds the amount which would normally be governed by the resistivity of the germanium.

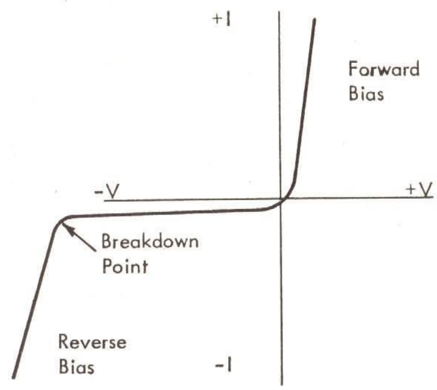
Notice in Figure 18-C that the potential hill has been reduced considerably from its reverse bias condition. However, a slight barrier remains because the junction contains some resistance and a slight voltage drop develops when current is passed. Gradual voltage drops are developed across the semiconductor but the bulk of the voltage is developed across the load resistor.

A high current transfer is possible when there is a minimum of recombination of minority carriers. In the example of the N-P diode, a high electron flow will result when there is a low concentration of holes in the P material; this is called electron injection or minority carrier injection. A predominance of hole flow would result if the N material contained fewer impurities than the P material.

Semi-conductor diodes develop only a few tenths of a volt drop and consequently exhibit a low impedance, when conducting in the forward direction. These devices should be protected with a current limiting resistor when passing current in the forward direction. A comparison is made between a forward biased semi-conductor, crystal and vacuum tube diode in Figure 18D.

From the discussion to this point, notice that the junction of two dissimilar types of semi-conductors possesses a definite rectifying peoperty. Its ability to pass current in only on e direction is an important characteristic of semi-conductor junctions.

A typical characteristic curve for a junction diode is shown in Figure 19.

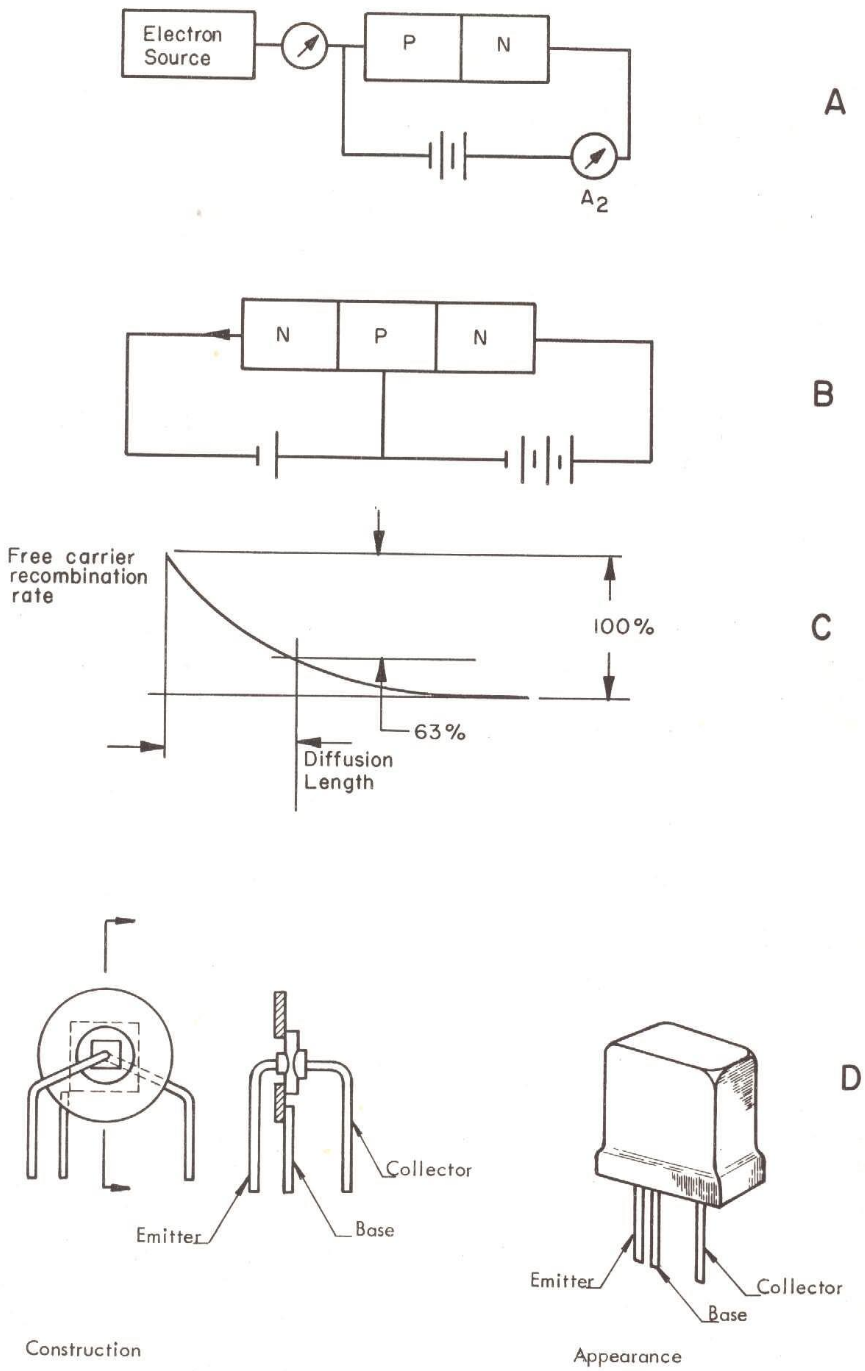


Diode Characteristic Curve

Figure 19

NPN JUNCTION TRANSISTOR (FIGURE 20)

The study of semi-conductor diodes provides a foundation on which an explanation of transistor action can be based. Expanding this diode theory provides a convenient and practical approach to the understanding of transistors. A diode characteristic curve shows that forward current is practically independent of forward bias voltage, and reverse current is essentially constant over a fairly wide range of reverse bias voltage. This condition makes it practically impossible to control the amount of forward bias current, but the characteristics of the diode could be altered if a means were devised for controlling the amount of reverse minority carrier current. Obviously, a change in temperature would reflect itself in a change in reverse current, but this action would be too slow to be of practical value in high speed circuits. What we're looking for is a method of controlling current flow; forward current appears to be uncontrollable but reverse current shows possibilities for control.



NPN Junction Transistor

Figure 20

A reverse biased P-N diode with a separate hypothetical electron source is shown in Figure 20A. Only the normal reverse diode current flows prior to the application of the external electron source. Remember this reverse current is thermally generated minority carrier current! Adding electrons to the P region can be considered to be an additional means of developing minority carriers. Some of these carriers may combine with the excess holes in the P region, but those that do not recombine come under the influence of the reverse bias voltage and are drawn across the P-N junction. By controlling the number of electrons entering the P material, we can control the amount of reverse bias current.

We have seen that a forward biased semi-conductor diode supplies electrons from an N region to a P region. If an N material is fused to the P side of an existing P-N diode, and connected with a forward bias, the resulting action develops an electron source which feeds into the P material. The three element structure is a junction transistor. The P material is called the base, the added N material electron source is called the emitter and the remaining N section is called the collector.

The amount of current passed by the forward biased diode is the sum of two components; electrons traveling N to P and holes traveling P to N. The ratio of these components is determined by the numbers of available electrons and holes at the junction. This in turn is controlled by the ratio of donors in the N region to acceptors in the P region. These impurity concentrations can be expressed in terms of resistivity; alloys with large numbers of free carriers have low resistivities. From basic Ohm's law, it can be concluded that the total current across the junction is a function of the voltage across the junction, and the resistivity of the junction. The net current is partially influenced by the concentration of majority carriers in each region. To obtain a minimum of recombination of electrons and holes in the base, the P type base is doped with a lower concentration of impurities than the emitter. This unequal concentration is represented by a typical ratio of emitter to base resistivities of one hundred to one. The various currents can be analyzed in Figure 20A. The increase in collector current, as noted at meter A2, is less than the increase in injected carriers as indicated by meter A1 because of the recombination and cancelling effects of electrons in the P material.

Minority carrier lifetime is affected by the recombination rate and the length of material the minority carrier must travel before recombination occurs. The measurement of this phenomenon is called "diffusion length" which is defined as the length of material at which the number of injected carriers reduces by 63% of its original value. A curve of free carriers vs length of material (Figure 20C) shows an exponential recombination rate that is similar to the discharge curve of a capacitor. In order to have a large percentage of the injected carriers cross the junction into the external circuit, the thickness of the base region must be considerably less than the diffusion length of the minority carriers. Percentages ranging upwards of 98% can be

realized when the base thickness is only two to three thousandths of an inch. The physical arrangement of the three transistor elements are shown in Figure 20D

PNP JUNCTION TRANSISTORS

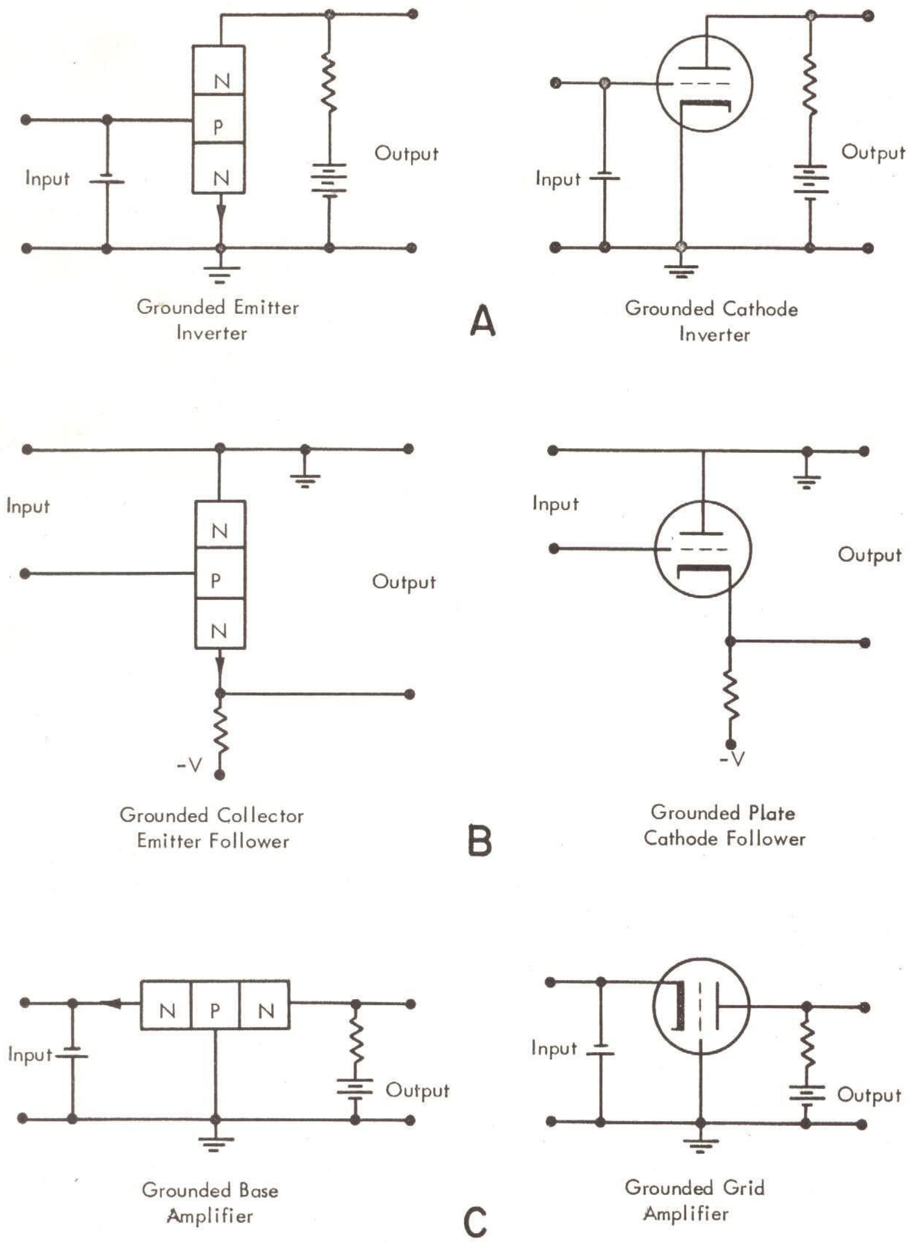
Reversing the impurity types in each section of the three element transistor results in a PNP unit. N type germanium is used for the base and indium dots are fused to the crystal to form emitter and collector areas. Its operation follows exactly the same theory as the NPN version except that the majority carriers from the emitter are holes. In the unbiased state, emitter holes diffuse into the base, and electrons from the base diffuse into the emitter until a potential is built up that retards any further diffusion between the two regions. A similar action takes place at the collector-base junction; diffusion takes place until further action is stopped by the P collector material becoming negative with respect to the base. This diffusion establishes a potential hill, high in the base region, that must be overcome before transistor action can take place.

PNP and NPN transistors are essentially complements of one another and very often the circuit designed for one can be used with the other merely by changing the polarity of the bias potentials. This has a definite advantage in the design of logical circuits and provides systems that are more direct than comparable vacuum tube circuits.

TRANSISTOR CONNECTIONS (Figure 21)

Transistors are used in electron circuits requiring two input and two output terminals. Because the transistor has only three terminals, one of these leads must be common to both the input and output circuit. Any one of the three leads can be used as the common terminal and it is customary to describe the mode of operation in terms of which terminal is common.

Direct analogies can be made between NPN transistor and vacuum tube circuits; therefore, the following illustrations refer only to the ~~P~~NPN variety. Similar circuits are available using PNP transistors. When an input signal is injected between the base and emitter and the output is taken between collector and emitter, the common element between the two circuits is the emitter. This is called a common or grounded emitter circuit and is shown with its vacuum tube counterpart in Figure 21 A. In both cases the circuit performs the function of an inverter. Figure 21 B shows the grounded collector transistor compared to the grounded plate tube circuit. Here the collector is the common element and the output is taken between the collector and emitter. This is called an emitter follower and performs a function similar to a conventional cathode follower. Figure 21 C demonstrates



Transistor and Tube Circuit Comparisons

Figure 21

a third possible arrangement that uses the base as the common terminal; this is compared to a vacuum tube grounded grid amplifier. Each configuration demonstrates a particular advantage; however, in all cases four basic rules apply to transistor circuits.

1. At the emitter-base junction, a potential hill builds up due to diffusion of majority carriers into the opposite region.

2. To insure that the transistor is cut off, the emitter-base junction is reverse biased with a voltage that increases the potential hill between the two regions.

3. To insure that the transistor is driven into conduction, the emitter-base junction is forward biased with a voltage that decreases the potential hill between the two regions. This presents a low impedence to the driving source and allows carriers to transfer from the emitter to the base.

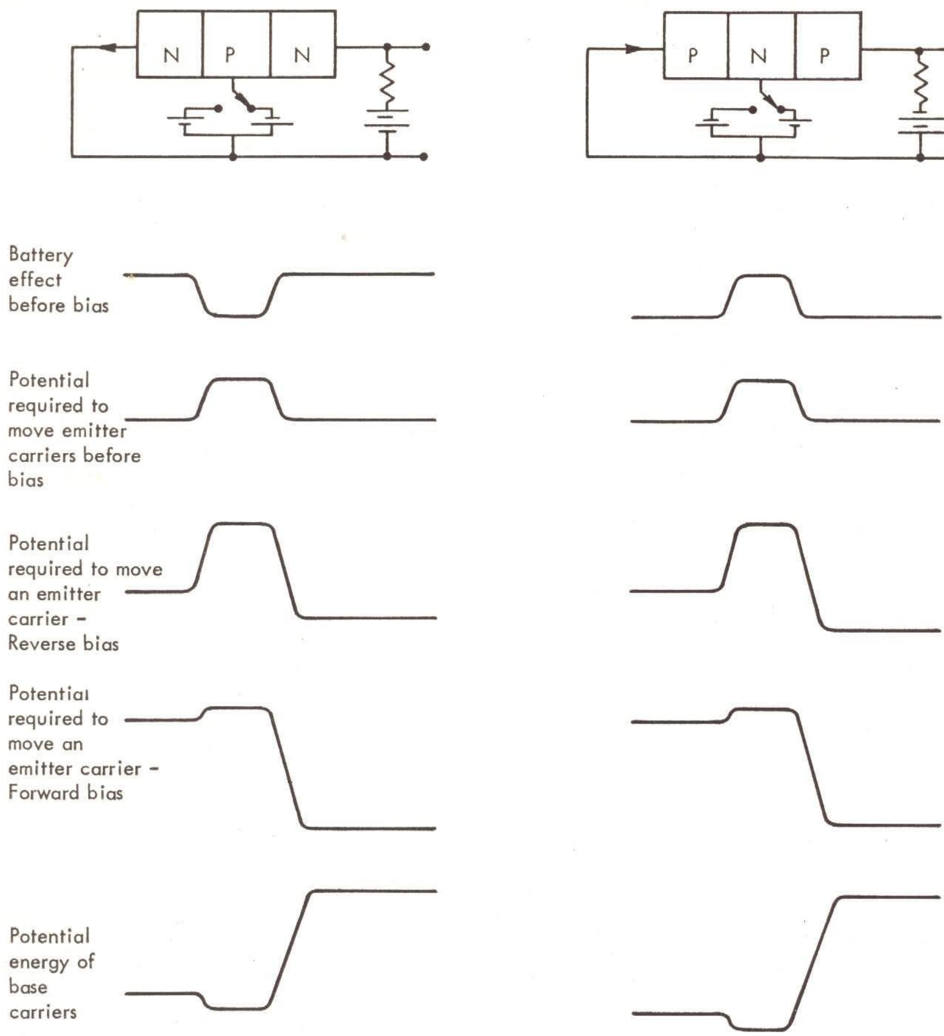
4. The collector is always biased in the high impedance or reverse direction to increase the potential hill between the base and collector. This aids in the transfer of emitter carriers through the base to the collector.

Reversed Biased Transistors. Figure 22

Transistors are operated in any of the standard modes used in tube circuits. They can be used as RF or AF amplifiers and are ideally suited for system logic design because they approximate a switch. When the unit is cut off its output impedance is a megohm or more; this value drops to less than fifty ohms when current is being carried. The transistor is basically a current device and is controlled by changing input current signals. Practical applications, such as machine systems, use voltage changes to indicate various conditions within the system. These voltages are converted to current changes at the input to the transistor and changed back to voltage changes at the output. An output signal is a linear reproduction of an input current rather than voltage change. The intermediate current ranges are not significant when transistors are used as a switch and are operated either at cut off or in saturation. Under these conditions, consider the input merely as a bias voltage between the base and emitter. It should be noted at this point that transistors need not be driven into saturation to be used as a switching device but this is a convenient approach to an introduction to transistor action.

Collector Cut-Off Current I_{cbo} or I_{co}

The NPN circuit of Figure 22A is shown with a switch controlling the base to emitter bias voltage. When the base is negative with respect to the



Potential Hills of Transistors

Figure 22

emitter, the potential hill is increased and majority carriers cannot cross the barrier. The same action occurs at the collector junction because it is also reverse biased. The previous discussion on collector bias indicates that only minority carrier current is able to cross the junction. This is the normal reverse current present when the transistor is cut off and is appropriately called I_{co} (cut off current). I_{co} increases rapidly with increases in temperature; it is measured with the emitter circuit opened. Measurable effects indicate that I_{co} increases ten times for each 30^{o} rise in junction temperature

Breakdown BV_{cbo}

Transistors are affected by reverse collector voltages in much the same way as diodes are affected by reverse voltages. Regardless of the current passed by the emitter circuit, increased collector voltages result in increased electric fields; these fields tend to pull electrons out of the valence band. The electrons are minority carriers in the P material and may dislodge other electrons due to atom bombardment; this also results in increased numbers of holes. The process is additive and large currents result with small increases in collector voltage after the critical (breakdown) voltage has been reached. A typical value of breakdown is 20 to 40 volts. Practical applications should not exceed 10 volts between the collector and base for reliable circuit operation.

Punch Through V_{pt}

The value of reverse voltage on the collector-base junction can be increased only until another critical point called "Punch Through" has been reached. Any increase in collector voltage beyond this point is reflected as an increase in emitter potential because the resistance of the base material is effectively lowered. Increased values of collector voltage draw donor electrons from the collector and leave behind bound positive charges. A migration of electrons causes the bound positive charges to appear near the collector-base junction. Excess electrons are given up by the battery so that they enter the P material of the base. Adding electrons to the base material cancels the effect of some of the holes developed by the impurity atoms. Cancelling holes effectively removes majority carriers near the collector junction; this has been defined as a depletion region because the region is being depleted of majority charge carriers. The base side of the junction develops a wider depletion area than the collector side because the base is less highly doped; therefore, has fewer majority carriers with which the electrons can combine. Increased voltages widen the depletion region to the point where the base is depleted of all its majority charge carriers. Further voltage increases merely add excess electrons to the P material and this effectively reduces the resistance of the base. The emitter, base, and collector each contain an excess number of electrons after this critical voltage is reached and the entire transistor acts as a low resistance device.

Increased collector potentials no longer develop a drop across the transistor; therefore, increased voltages appear as increases in the potential of the emitter. The point at which this action starts is called the punch through voltage and occurs at about 15 volts.

Forward Biased Transistor

Reversing the switches in Figure 21 forward biases both the NPN and PNP transistors. In each case the emitter-base potential hill is lowered by a voltage that provides an attractive force to the majority carriers in the emitter. It is convenient to remember that transistors are driven into conduction with a voltage polarity equal to the type of base material; e.g. NPN units turn on with a positive voltage and PNP units turn on with a negative voltage.

Forward biasing the emitter-base junction causes emitter carriers to travel into the base to become minority carriers. The base-collector junction is reverse biased for majority carriers but forward biased for minority carriers; therefore, the emitter carriers that enter the base are drawn into the collector circuit. In the transfer from the emitter through the base, some of the carriers may combine with the majority carriers of the base but because the base thickness is very small, the impurity concentration of the base is low, and the attractive force of the collector is high, the net result is a high percentage of emitter carriers entering the collector region. The total collector current is the normal I_{co} plus the amount of emitter current that does not recombine in the base to form base current. The transistor action is summarized as a means of controlling the amount of reverse current through a reverse biased diode.

In most written work on transistors, conventional current flow (+ to —) is used to describe both NPN and PNP units. The direction of the emitter arrow indicates this current flow. A more basic, yet clearer approach can be made if emitter majority carriers are considered regardless of the type of carrier involved. NPN units have electron flow from emitter, through the base to the collector. PNP use the same path except that holes are the majority carriers from the emitter. This concept results in identical paths for both units but it must be understood that the type of carrier changes when shifting from one type unit to the other. Two schools of thought are available and are in common use. Electron flow from - to + is used in NPN circuits; conventional current flow from + to - is in the same direction as hole flow and is adapted to PNP circuits.

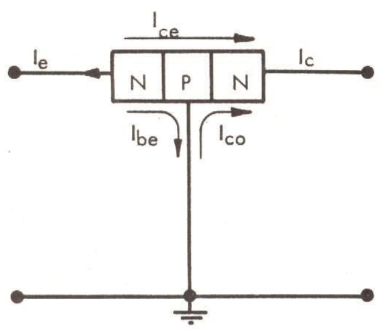
Grounded Base Circuit

The typical grounded base connection (Figure 23A) shows the various electron paths used in the input and output circuits. Collector current consists only of I_{co} before an input current source is supplied; this is indicated on the volt-ampere characteristic (Vc-Ic) curve as Ie = O. Increasing the emitter source to 1 ma increases the collector current to 1 ma, plus the normal I_{co}, minus the amount of current that recombines in the base. Further increases in emitter current results in corresponding increases in collector current; however, the change in output current is always less than the input change because of the base recombination. A grounded base junction transistor actually delivers less current at its output than is received at its input; however, a power gain is developed because of the difference in impedance levels between the input and output circuits. The emitter-base is a forward biased diode with an impedance near fifty ohms. The output is a reverse biased diode with an impedance of about a megohm.

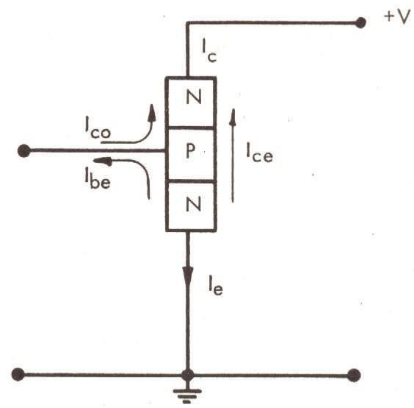
Characteristic Curves

The Vc-Ic characteristic curves best describe the transistor action of the grounded base circuit. Collector currents are measured at various collector voltages with the emitter current held constant. A family of curves develops when several values of emitter current are plotted. Each value of input current develops an essentially constant value of collector current throughout the range of the collector reverse bias voltage. The curves show that a slightly forward biased collector junction is needed to reduce the collector current to zero. This is necessary because the forward bias of the emitter-base junction causes many of the emitter carriers attain a sufficient velocity to cross the collector junction, regardless of the magnitude of the aiding field. As the collector reverse bias is reduced below zero, some current flows from the collector to base that subtracts from the normal emitter current reaching the collector. The small forward collector-base voltage retards emitter carriers and starts a flow of collector to base carriers. When the collector to base, and emitter to collector currents cancel, the collector current drops sharply to zero.

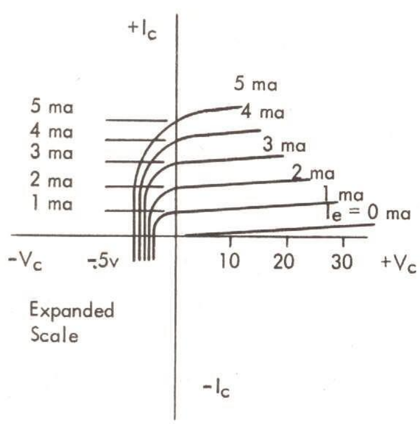
This circuit is used for impedance matching, voltage amplification and power gain.



Electron
Flow Paths

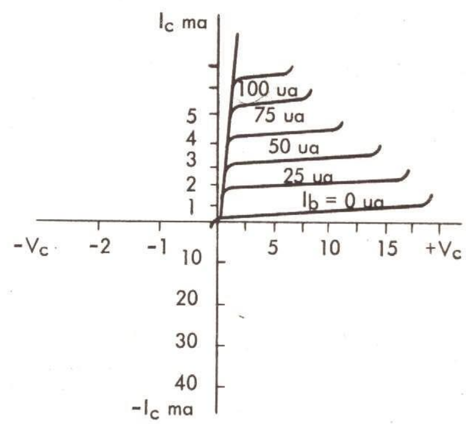


Electron
Flow Paths



A

Grounded Base NPN
V_c-I_c Characteristics



B

Grounded Emitter NPN
V_c-I_c Characteristics

Figure 23

Alpha

Current gain is represented by the Greek symbol alpha (α) and is defined as the ratio of change in collector current to the change in emitter current when the collector voltage is held constant. Because some recombination occurs in the base, collector current changes are less than emitter current changes and the ratio is always less than one for junction transistors. Typical values of α are from .95 to .99.

Grounded Emitter and Grounded Collector Circuits

Base input circuits are affected by the low impedance of the emitter-base diode as well as the high impedance of the collector-base diode. This presents a higher impedance to the driving source than the previously discussed grounded base circuit. Input signals are applied to forward bias the emitter-base diode and a large number of carriers cross the junction. From the discussion of a forward biased diode it was noted that only a small influence was necessary to initiate a scattering process and a cumulative flow of carriers. Small current signals of majority carriers into the base induces a large flow of emitter carriers to cross the junction. Again emitter carriers come under the influence of the collector potential and are drawn across the collector junction.

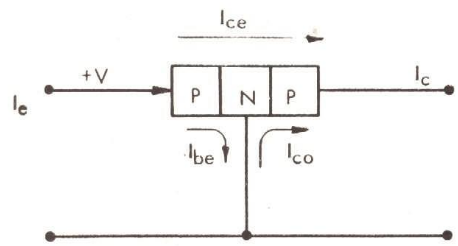
A typical grounded emitter connection (Figure 23 B) shows the various electron paths involved in the input and output circuits. The Vc-Ic curves show that when no base carriers are supplied, only I_{co}' flows as collector current. I_{co}' is an amplified cutoff current equal to $I_{co} \div 1 - \alpha$. When 25ua is injected into the base, about 2 ma flows in the collector circuit. This connection provides an appreciable current gain between the input and output circuits.

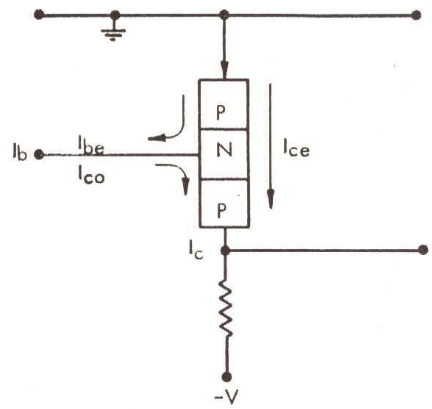
Alpha Prime

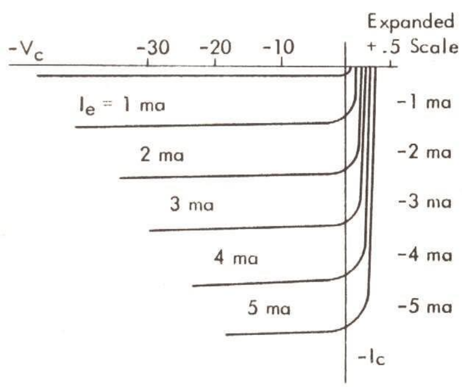
(α') is the symbol used to designate the current gain between the collector and base. It is definied as the ratio of the change in collector current to the change in base current with a constant collector voltage. This is an important transistor parameter and is dependent upon the construction of the transistor as well as the circuit in which it is used. Typical values range from 50 to 150 (Alpa prime is also commonly refered to as Beta)

PNP Grounded Base - Figure 24

Connections for PNP transistors involve only the changing of collector and bias potentials from equivalent NPN circuits (Figure 24 A). Holes make up the collector current and are in a direction leaving the transistor; therefore, negative. Collector potentials are also negative and the Vc-Ic curves are developed in the third graphical quadrant. NPN grounded base theory is duplicated in this connection except that the majority carriers from the emitter are holes. Alpha remains less than one.

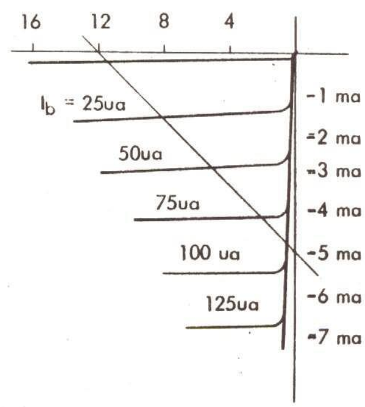






A

Grounded Base



B

Grounded Emitter

PNP Collector V_c-I_c Characteristics

Figure 24

PNP Grounded Emitter

Again NPN theory can be adapted to the PNP variety and analyzed as hole carriers traveling through the unit in the directions indicated in Figure 24B

Load Lines

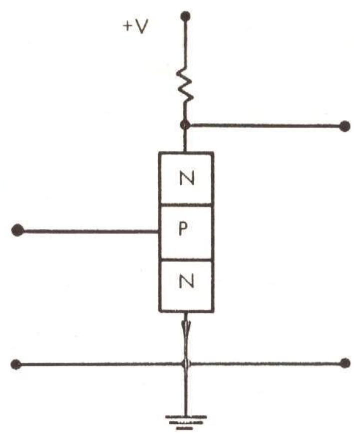
The Vc-Ic characteristic is similar to the previously discussed version except that a typical load line has been added to this diagram. Terminating points are established on the voltage axis at the open circuit voltage across the transistor and on the current axis at the short circuit current through the transistor. A cut off transistor has only Ico' flowing; this intersects the load line and the drop across the transistor is measured horizontally from the zero reference point. Almost the entire terminal voltage appears across the transistor because only a small current is flowing; this indicates that the unit has a high impedance when cut off. An input signal of 25ua develops a collector current near 1.5 ma; the voltage across the transistor drops to 8v and 4v appears across the load resistor. Increasing the input signal to 100 ua develops 5 ma of collector current. This is the saturation current for the transistor and is limited by the power dissipation ability of the unit. The voltage across the transistor drops to a very low value; this is usually less than .3v. The high current and low voltage drop indicates that the transistor has a very low impedance when in saturation.

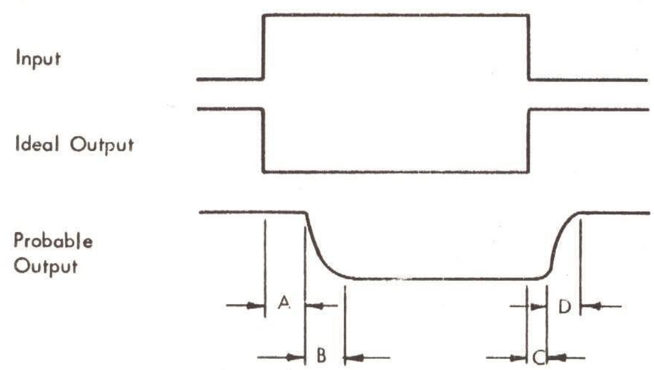
Transient Effects - Figure 24

Analyzing the grounded emitter connection of Figure 24 shows that the circuit is functioning as an inverter. Normally the transistor is cut off and output is near the positive collector return voltage. A positive input signal drives the unit into conduction, only a small voltage drop develops across the transistor and the output approaches the emitter return voltage. Ideally a positive square wave input should produce a negative square wave output. The ideal square wave is not achieved because of two transistor parameters that effect the output signal; one is transit time and the other is minority carrier storage in the base. An NPN unit is discussed; however, the same theory applies to PNP transistors.

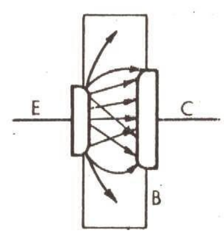
Turn on Delay

A finite difference in time results when an electron is passed by various types of materials. N type semiconductors contain an excess number of electrons; adding an electron to one side of this material immediately results in a different electron leaving the opposite side so that equilibrium can be maintained. It is not necessary that the electron travel the full width of the material before an electron leaves the opposite side. One electron is just as effective a carrier as another; therefore, the net transfer can be considered to have a very small transit time.





A. - Transistor Turn On Delay
B. - Turn On Transition
C. - Transistor Turn Off Delay
D. - Turn Off Transition



Transistor Transient Effects

Figure 25

Adding electrons to P type semiconductors produces a slightly different effect. This material contains few, if any, free electrons. When an electron is added to one side, a finite interval of time exists before it arrives at the opposite side. In NPN transistors a delay results because electrons must cross the P type base region. Electrons leaving the emitter arrive at the collector some interval of time later as represented by "A" in Figure 24. This is called the "turn on delay time".

Turn Off Delay

When a negative input signal reverse biases the emitter-base diode; electrons in transit across the base region continue to flow to the collector circuit. Some of the electrons may combine with holes for short periods of time during the transfer across the P material; this is a "minority carrier storage" effect. Both of these actions effect the interval of time between the last electrons leaving the emitter and the last electrons arriving at the collector. This is represented as time "C" and is called the "turn off delay time".

Turn On Transition.

Unfortunately electrons do not all leave the emitter with the same velocity; consequently they arrive at the collector at different times. In addition to this, the electrons may take any of several random paths between the emitter and collector. The combined effect is that the fastest and most direct electrons arrive at the collector first, the slower electrons arrive a short time later and finally the slowest electrons that took the most indirect path arrive at the collector last. The net result is a gradual build-up of carriers reaching the collector, and a rather steady arrival rate when the transistor is in full conduction. This effect may show up as a sloping leading edge of the output signal and is defined as the "turn on transition time".

Turn Off Transition

The difference in velocities and path length also affect the arrival time of the last electrons to reach the collector. A gradual tapering off of collector carriers results when the transistor is cut off. This is the "turn off transition time".

Frequency Cutoff - Figure 26

The random speeds and paths of electrons leaving the emitter develop a spread in arrival times at the collector. This spread in arrival time is called the "dispersion interval". An emitter pulse having a width of one unit of time may be delayed five units of time and then spread itself over eight units of time at the collector. The total amount of carriers may be the same but the time interval is quite different. The curve of collector arrival time is approximated by two straight dashed lines in Figure 25.

A

Collector arrival rate of emitter carriers

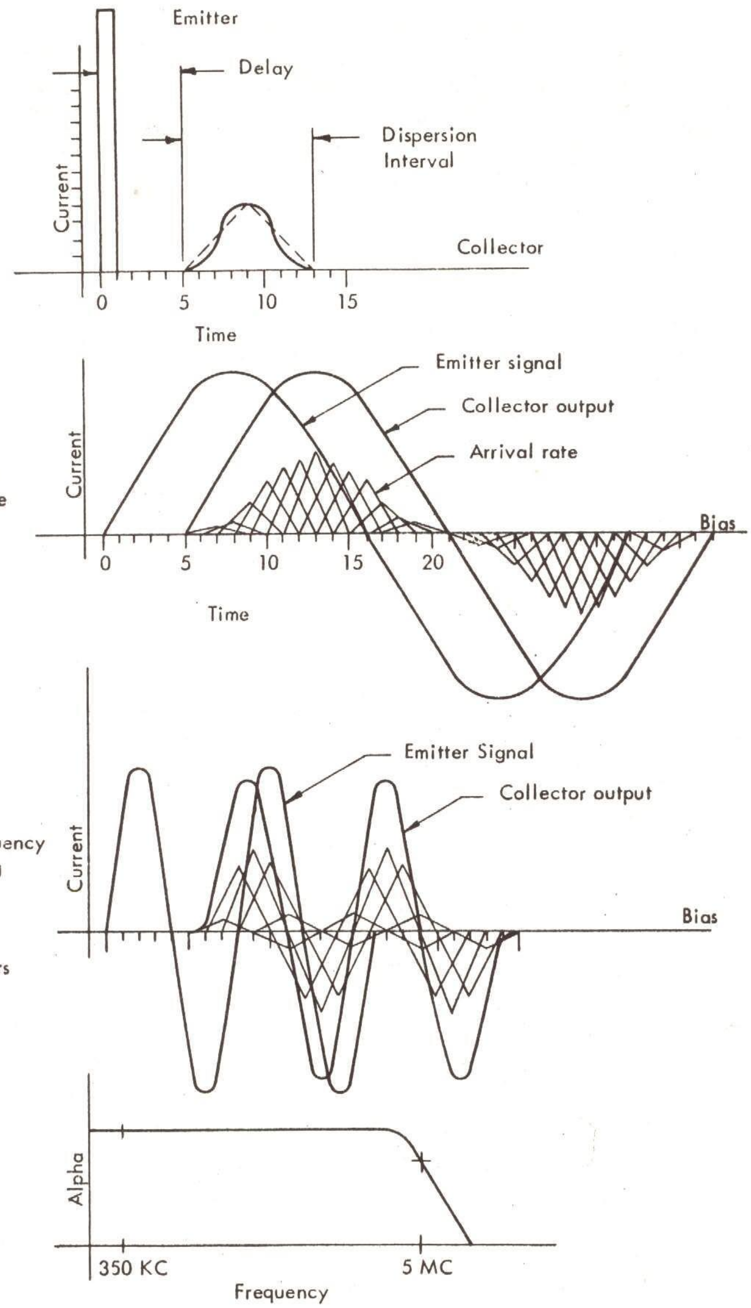
B

Composite arrival rate of emitter carriers.

C

High frequency cancelling effects of collector current components

D



Transient Action
Figure 26

Emitter signals having a pulse time that is long compared to the time of the dispersion interval display no unwanted effects in the carrier travel through the transistor. This is shown in Figure 25. where the total collector current is made up of components of several dispersion intervals. Throughout the sine wave cycle, components add to each other and the current output is equal to alpha times the emitter input signal.

When emitter signals are short compared to the time of the dispersion interval, collector current does not follow the emitter current. The width of the dispersion interval causes some of the maximum amplitude current components to flow during the time of minimum emitter currents. In the same way, minimum amplitude currents flow during the time of emitter maximum amplitude. Total collector current is the net sum of all components; therefore, if some components cancel others, the net result is a reduction in collector current. A measure of this property is the frequency cut off point (f_{co}). It is also called the alpha cut off point and is defined as the frequency at which alpha reduces to 70.7% of its low frequency current gain. An arbitrary low frequency reference point that is becoming standard is 350 kc. Current gain is measured at this frequency and remains rather constant as frequency is increased up to some upper limit. Alpha decreases rapidly beyond this point until the transistor no longer functions efficiently. A typical point at which the current gain reduces to 70.7% (3 db) is 5 megacycles.

Alpha cut off can be increased by reducing the width of the base region but this also reduces the punch through voltage; therefore, a compromise is made for the two parameters.

IBM Transistor Specifications and Types

Variations in materials, geometry, and construction result in transistors with widely varying parameters. Transistor parameters are characterized to fall within given limits and are assigned a type number. A two digit numerical series with 01 is assigned to all PNP transistors; a similar two digit series starting with 51 is used for NPN varities Transistor Parameters can be summarized as follows:

Collector Cut-Off Current, I_{cbo}

The emitter is open circuited and a reverse voltage is applied to the collector-base junction. Maximum collector current is specified in micro-amperes.

Collector Cut-Off Current with Reverse Biased Emitter, I_{cbo}

With the base common, and the emitter and collector junctions both reverse biased the reverse collector current cannot exceed a specified limit. This reverse biased transistor current should be no more than I_{cbo}.

Emitter Cut-Off Current, I_{ebo}

The collector is open circuited and a reverse voltage is applied to the emitter-base junction. Maximum emitter current is specified in micro-amperes.

Collector Breakdown Voltage, BV_{cbo}

Collector breakdown voltage is determined by passing a controlled amount of reverse current through the collector-base junction and meas-using the collector voltage drop. The emitter is open circuited for this test.

Emitter Breakdown Voltage, BV_{ebo}

Emitter breakdown voltage is determined by passing a controlled amount of reverse current through the emitter-base junction and measuring the emitter voltage drop. The collector is open circuited for this test.

Punch Through Voltage, V_{pt}

A specified reverse collector to base voltage is applied and the emitter to base voltage should fall within an established upper limit.

Current Ratio, Common Emitter, α_{FE}

Collector current is measured at specified values of base current and collector voltage to establish a minimum current ratio for the transistor.

Collector to Emitter Saturation Voltage, V_{ce}

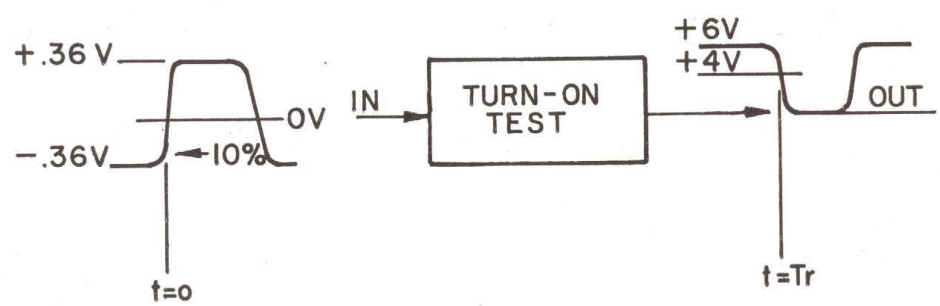
Sufficient collector voltage and base current is supplied to drive the transistor into saturation. At this point the impedance of the transistor becomes very low and is measured as a low voltage drop across the emitter-collector terminals. A maximum voltage, in the order of tenths of a volt, is specified for this parameter.

Base to Emitter Saturation Voltage, V_{be}.

This measurement is made under the same conditions as the V_{ce} test.

Common Emitter Turn-On Time, T_{on}.

Turn on time is measured as the interval of time between the 10% point of an input pulse and the point at which the output pulse reaches a specified voltage level. A particular circuit and condition is used to determine this turn-on parameter.



Alpha Cut Off Frequency, $f\alpha_c$

Alpha cut off frequency is the frequency is the frequency at which the magnitude of the small signal common base current ratio (α_{fe}) has fallen to .707 of its low frequency value.

K Factor

The ratio of the temperature rise at the junction to the power dissipated by the transistor is called the K factor. Low K factor units dissipate more heat than do units with higher K factors therefore have smaller effects on I_{cbo}. Typical IBM transistors have a K factor range between .2 and .8.